# Cdx2 maintains stem-cell identity in the mouse trophoblast lineage by preventing default differentiation as trophoblast giant cells

Bozon, Kayleigh Amanda

*Awarding institution:*
King's College London

# Cdx2 maintains stem-cell identity in the mouse trophoblast lineage by preventing default differentiation as trophoblast giant cells

**Kayleigh Amanda Bozon**

King's College London
and
The Francis Crick Institute
PhD Supervisor: Jim Smith

A thesis submitted for the degree of
Doctor of Philosophy
King's College London
September 6, 2019

## Declaration

I Kayleigh Amanda Bozon confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

# Abstract

The transcription factor *Cdx2* is required for the formation of several developmental lineages in the mouse embryo, and is thought to play different roles within the same lineage at different stages. The role of *Cdx2* in establishing trophoblast fate has been extensively studied, but the reason for its continued expression in established trophoblast cells is less well understood.

To explore this question, I generated ATAC-seq and RNA-seq libraries from wild-type and *Cdx2* knockdown trophoblast stem cells (TSCs) *in vitro*. ATAC-seq experiments show, as expected, that loss of *Cdx2* causes decreased accessibility at CDX2 consensus binding sites. However, of the regions showing differential chromatin accessibility in *Cdx2* knockdown cells, two thirds have increased accessibility. These sites are enriched for the TFAP2C consensus motif and footprint; *Tfap2c* is a TSC marker that plays a part in trophoblast differentiation by driving genome-wide increases in accessibility.

Gene ontology analysis of sites with increased accessibility in *Cdx2* knockdown cells suggests that these regulatory regions are associated with trophoblast giant cell (TGC) differentiation. Consistent with this observation, continued *Cdx2* knockdown drives homogenous differentiation of trophoblast stem cells into trophoblast giant cells within days. Similarly, homozygous *Cdx2* knockout TSCs are unstable and spontaneously differentiate into TGCs, although in this context cells appear to retain some ability to compensate.

TSC differentiation can also be initiated *in vitro* by growth factor withdrawal, albeit in a heterogeneous manner. However, the sites of increased chromatin accessibility observed under these circumstances differ from those observed in *Cdx2* knockdown cells. Moreover, although RNA-seq analysis shows that 87% of genes whose expression changes in *Cdx2* knockdown cells are also mis-regulated during conventional differentiation, 31% of these transcripts change in opposite directions.

This work suggests that *Cdx2* normally maintains ‘stemness’ in TSCs by preventing them from differentiating directly and homogeneously into TGCs. This gatekeeping role of *Cdx2* differs from its earlier function in repressing pluripotency genes and suggests that *Cdx2* is repurposed to play different roles within the trophoblast lineage. Furthermore, our work may finally provide a plausible mechanism by which half of non-progenitor derived parietal TGCs are generated.

# Acknowledgements

I am grateful to so many people for their input and support over the past four years that it is difficult to know where to start. Firstly, I would like to thank my supervisor Jim for providing me with this opportunity, for his supervision and for allowing me the freedom to explore my scientific whims.

I am completely indebted to every member of the Smith Lab, past and present. Thank you, Andreia, for your input into my project and our discussions over the years. I am also incredibly thankful to Fay for her fast chapter turn-arounds and her patience with my many questions a day for several years. Many thanks to Rita and George for teaching me basic bioinformatics and for their input into experimental design and techniques. Thank you, Camille and Greg, for being there and teaching me at the start.

This thesis would not have been possible without the resources available and my colleagues at The Francis Crick Institute; and thank you for funding my work. I have to say a huge thank you to Daniel for teaching me all things trophoblast stem cell and for continually sharing your time, input and resources with me. I'm grateful to the Advanced Sequencing Facility for preparing and sequencing my libraries and I'd like to extend a special thank you to Harshil for his analysis of the sequencing data.

Thank you to my Thesis Committee, Kathy Niakan, Jack Price and Vivian Li, for your different perspectives and input into making this thesis what it is.

Thank you to Parveen and Sonia, for putting up with me despite my inability to respond to messages. I have to say a special thank you to Enoch for teaching me to kill the closest crocodile first and for continually checking in, no matter what. Thank you to Sophie for making space for me in your home, as well as your and Sang Eun's support over the past few years.

Whilst writing this thesis, I was given a cruel reminder of why placental research is so important. Although her life was short, my cousin Harper-Rose was tiny, perfect

and strong. I wish you were here rather than me writing these words. Thank you for giving me some perspective and for reminding me that we still have much to learn and many little lives to protect.

Words will never be enough to describe how invaluable the support and love of my family has been over the past 26 years, not just the past four. Thank you, Liam and Nicole, for reminding me to let my hair down. Thank you, Blake, for putting up with my ‘work-work’ life balance and for our many adventures; here’s to our next one on the other side of the world. Finally, my biggest thank you is for my parents, for your unconditional love and for giving me everything I could have ever wanted. None of you understand much of what I do, but that is all the more reason to dedicate this thesis to you.

“Don’t adventures ever have an end? I suppose not. Someone else always has to carry on the story.”
*J.R.R. Tolkien, The Fellowship of the Ring*

# Table of Contents

# Table of figures

# List of tables

# Abbreviations

| | |
|---|---|
| 3’ | 3-prime |
| 4C | Circular chromosomal conformation capture |
| 5’ | 5-prime |
| ATAC-seq | Assay for Transposase Accessible Chromatin with high-throughput sequencing |
| bHLH | Basic helix loop helix |
| C1 | Cluster 1 |
| C2 | Cluster 2 |
| C3 | Cluster 3 |
| cDNA | Complimentary DNA |
| ChIP-seq | Chromatin immunoprecipitation with high-throughput sequencing |
| CRISPR | Clustered regularly interspaced short palindromic repeats |
| c-TGC | Canal trophoblast giant cell |
| Ch-TGC | Channel trophoblast giant cell |
| d2DARs | Sites of gained accessibility two days after growth factor withdrawal |
| DAPI | 4’,6-diamidino-2-phenylindole |
| DAPI-A | 4’,6-diamidino-2-phenylindole area |
| DAPI-H | 4’,6-diamidino-2-phenylindole height |
| DD1 | Differentiation by growth factor withdrawal day one |
| DD3 | Differentiation by growth factor withdrawal day three |
| EPC | Ectoplacental cone |
| ESC | Embryonic stem cell |
| FACS | Fluorescence activated cell sorting |
| FSC-A | Forward scatter area |
| ExE | Extraembryonic ectoderm |
| ETBS | Embryo-derived trophoblast stem cell CDX2 binding site |
| FPKM | Fragments per kilobase of exon model per million reads mapped |
| GA | Gained accessibility sites in *Cdx2* knockdown |
| gDNA | Genomic DNA |
| GO | Gene ontology |
| GREAT | Genomic Regions Enrichment of Annotations Tool |

| | |
|---|---|
| GRN | Gene regulatory network |
| HTH | Helix-turn-helix |
| HTS | High-throughput screening |
| ICM | Inner cell mass |
| IGV | Integrative genomics viewer |
| IF | Immunofluorescence |
| iTSC | Induced trophoblast stem cell |
| KD | *Cdx2* knockdown |
| LA | Lost accessibility sites in *Cdx2* knockdown |
| MACS | Model-based Analysis of ChIP-Seq |
| mESC | Mouse embryonic stem cell |
| MEF | Mouse embryonic fibroblast |
| MGI | Mouse genome informatics |
| mLA | Minus sites of lost accessibility in *Cdx2* knockdown |
| NMD | Nonsense mediated decay |
| PCA | Principal component analysis |
| PCR | Polymerase chain reaction |
| p-TGC | Parietal trophoblast giant cell |
| qRT-PCR | Quantitative reverse transcription polymerase chain reaction |
| RNA-seq | RNA sequencing |
| s-TGC | Sinusoidal trophoblast giant cell |
| SD | Standard deviation |
| shRNA | Short hairpin RNA |
| SCR | Scramble (mCherry positive) short hairpin RNA |
| SCR NEG | Scramble mCherry negative short hairpin RNA |
| SEM | Standard error of the mean |
| SPRI | Solid Phase Reversible Immobilisation |
| SpT | Spongiotrophoblast |
| SSC-A | Side scatter area |
| Syn | Syncytiotrophoblast |
| Syn-I | Syncytiotrophoblast-I |
| Syn-II | Syncytiotrophoblast-II |
| SpA-TGC | Spinal artery trophoblast giant cells |
| TE | Trophectoderm |

| | |
|---|---|
| TEE | Trophectoderm-specific enhancer |
| TF | Transcription factor |
| TGC | Trophoblast giant cell |
| TPM | Transcripts per million |
| TSC-like | Trophoblast stem cell like cell (derived by transdifferentiation) |
| TSS | Transcription start site |
| TSC | Trophoblast stem cell |
| UT | Untransfected |

# Chapter 1. Introduction

The placenta is required for healthy foetal development. Defects in the placenta result in its dysfunction and the reduced transfer of nutrients to the foetus. As a result, placental insufficiency causes many developmental defects *in utero* and has also been shown to pre-dispose individuals to late-onset disease (reviewed in Gagnon, 2003). Recent work by Perez-Garcia et al (2018) has shown that placenta insufficiency in mouse knockout models causes embryonic developmental defects directly and can exacerbate abnormal foetus development (Perez-Garcia et al., 2018). Loss of many placental genes results in embryonic death, but this can be rescued by providing a normal placenta (Perez-Garcia et al., 2018, Guillemot et al., 1994, Auman et al., 2002). For example, knockout of Caudal related 2 (*Cdx2)*, a transcription factor known to by required for healthy placenta development, is embryonic lethal at embryonic day 4.5 (E4.5) (Strumpf et al., 2005). Tetraploid complementation assays, where *Cdx2*-null embryos are provided with a wildtype placenta, rescues early embryonic lethality (Chawengsaksophak et al., 2004). In the first trophoblast (placenta) cells, the trophectoderm, CDX2 is required to antagonise pluripotency genes, silencing them to promote placental development (Strumpf et al., 2005, Niwa et al., 2005). However, *Cdx2* is expressed in trophoblast stem cell populations of the placenta long after this role becomes obsolete. What role *Cdx2* plays in trophoblast stems cells after lineage commitment remains unclear.

## 1.1 Early mammalian embryogenesis: making a blastocyst

The self-organising mouse embryo is a model system used to understand early development and lineage commitment. However, mouse and human embryonic development differs both in timing and in the underlying transcriptional network. Here, I briefly outline what is known about mouse blastocyst formation and some key differences between early mouse and human development.

### 1.1.1 The first lineage decision: trophectoderm or inner cell mass?

After fertilisation, the zygote undergoes successive cell divisions (Chazaud and Yamanaka, 2016). Historically, all cells of the early morula embryo were thought to be identical, but these have since been shown to have a bias for different developmental lineages as early as the 4-cell stage (Kelly, 1977, Fujimori et al., 2003, Tabansky et al., 2013). At the 8-cell stage, tension increases at the embryo surface, blastomeres form adherens junctions with one another and filipodia form over adjacent cells (De Vries et al., 2004, Fierro-Gonzalez et al., 2013, Maitre et al., 2015). This process results in a morphological change known as compaction. During compaction, embryos become polarised and the outermost cells accumulate the cell polarity pathway proteins, PAR3, PAR6 and aPKC, F-actin and microvilli at the apical surface (Yamanaka et al., 2006, Plusa et al., 2005, Alarcon, 2010). Conversely, another polarity protein, PAR1, accumulates basolaterally at the surfaces of cells with high cell-cell contacts (Vinot et al., 2005). Combined, these events lead to the formation of outer polarised cells and inner non-polarised cells.

Compacted and polarised 8-cell embryos undergo two further rounds of cell divisions (Chazaud and Yamanaka, 2016). Lineage commitment begins at the 16-cell stage and is complete by the 32-cell stage, when embryos undergo cavitation to form the blastocyst structure (Posfai et al., 2017). During this period, cells commit to one of two lineages: the trophectoderm (TE) and the inner cell mass (ICM) (Chazaud and Yamanaka, 2016). The outermost polarised cells commit to the TE, which generates the foetal proportion of the placenta, and the inner non-polarised cells commit to the ICM, which generates all other cell types (Chazaud and Yamanaka, 2016).

Commitment to the TE and the ICM is controlled by the differential activation of the Hippo signalling pathway in outer and inner cells (Nishioka et al., 2009, Cockburn et al., 2013, Hirate et al., 2013, Leung and Zernicka-Goetz, 2013). Briefly, within inner cells, phosphorylated AMOT localises to cell-cell adherens junctions, where it forms an active complex with NF2 and LATS1/2 (Hirate et al., 2013, Cockburn et al., 2013). Active Hippo signalling phosphorylates YAP, sequestering it to the cytoplasm and driving inner cells to acquire an ICM fate (Hirate et al., 2013, Leung and Zernicka-Goetz, 2013). In outer cells, AMOT is localised to the apical surface of outer cells by polarity proteins, preventing binding to LATS1/2 and turning Hippo signalling off. This enables YAP to translocate to the nucleus and act as a co-factor for TEAD4 (Nishioka et al., 2009, Hirate et al., 2013, Leung and Zernicka-Goetz, 2013). Recent work has shown that aberrant mid-range Hippo signalling in outer cells results in the ectopic expression of the ICM marker SOX2. As a result, such cells are either internalised and become inner cells or undergo apoptosis (Frum et al., 2018).

TEAD4 and the co-activator YAP activate the trophectoderm program in outer cells, inducing the expression of both *Cdx2* and *Gata3* (Nishioka et al., 2009, Ralston et al., 2010). Without *Tead4*, outer cells cannot commit to the trophoblast lineage, and embryos arrest after compaction (Nishioka et al., 2008, Yagi et al., 2007). *Cdx2*-null embryos can form blastocoel cavities, albeit briefly, and express some TE markers, indicating that *Cdx2* is not required for lineage segregation (Strumpf et al., 2005, Wu et al., 2010). Instead, CDX2 antagonises and represses the expression of core pluripotency markers in the trophectoderm (Strumpf et al., 2005, Niwa et al., 2005). As a result, *Cdx2*-null embryos retain *Oct4* and *Nanog* expression in the TE (Strumpf et al., 2005). Similarly, *Oct4*-null embryos express TE marker genes in the ICM (Nichols et al., 1998). Therefore, this antagonistic relationship between *Cdx2* and *Oct4* is critical to reinforce commitment to the TE or the ICM.

### 1.1.2 Trophoblast lineage segregation in human embryos

Differences in early mouse and human blastocyst development are numerous and have been reviewed in depth (Wamaitha and Niakan, 2018). The timing of developmental stages, including embryonic genome activation (EGA) and compaction, differs between human and mouse (Flach et al., 1982, Hamatani et al.,

2004, Braude et al., 1988)(Nikas et al., 1996, Steptoe et al., 1971, Edwards et al., 1981, Wamaitha and Niakan, 2018). Another critical difference between early mouse and human development is the timing of lineage segregation. In the mouse, segregation of the TE and ICM predates the decision in the ICM to form either the yolk sac primitive endoderm (PrE) or the embryo proper forming epiblast (Epi) lineages (reviewed in (Rossant, 2015) and (Chazaud and Yamanaka, 2016)). In contrast, formation of these lineages may be concurrent in humans, with all cells of the expanded blastocysts able to make TE, PrE and Epi cells (Petropoulos et al., 2016, De Paepe et al., 2013).

Additionally, the expression of known key lineage markers differs between human and mouse development. Whilst orthologs of some mouse TE markers are also expressed in human TE, such as *GATA3*, others are delayed or absent, such as *CDX2* (Blakeley et al., 2015). *Cdx2* is expressed before blastocyst formation in the mouse (Strumpf et al., 2005). In humans, *CDX2* is only expressed in TE cells post-cavitation at the 35-cell stage and is co-expressed with *OCT4* (Chen et al., 2009a, Niakan and Eggan, 2013). This is comparable to the timing of *CDX2* expression in cow, pig and rabbit embryos (Kuijk et al., 2008, Kobolak et al., 2009, Berg et al., 2011, Bou et al., 2017). Differences in *CDX2/Cdx2* expression between human and mouse TE may align with differences in timing of lineage commitment (Rossant, 2015, Petropoulos et al., 2016). In humans, where lineage commitment occurs later, *CDX2* and its downstream effectors may still be critical for the repression of *OCT4* expression (Niakan and Eggan, 2013).

## 1.2 Placental development and the trophoblast lineages

The placenta is a critical interface between mother and foetus, enabling exchange of nutrients, waste products and secreting factors to maintain pregnancy (Rossant and Cross, 2001). Amniotes, including mammals, birds and reptiles, form four common extraembryonic membranes: the amnion, chorion, allantois and yolk sac. Chorio-allantoic placentas of mammals are formed from both embryonic-derived and maternal-derived tissues. The epithelial proportion of chorio-allantoic placentae is derived from the trophectoderm (Mess and Ferner, 2010, Ferner and Mess, 2011).

### 1.2.1 Mouse placental development

The mouse has been used extensively as a model for placental development. There are multiple review articles that summarise our knowledge of mouse placental development to date (Rossant and Cross, 2001, Simmons and Cross, 2005, Watson and Cross, 2005, Woods et al., 2018, Rai and Cross, 2014). In this section, I will provide a brief overview of placental development and its specialised cell types. Figure 1.1b outlines the transition between different cell types and known marker genes during placental development.

#### *1.2.1.1 From one cell layer to functional placenta*

The fate of cells in the trophectoderm is dependent on their position. Those that directly overlay the ICM, the polar trophectoderm, receive FGF4 signalling and continue to proliferate (Rossant and Cross, 2001). The mural trophectoderm, which is not in contact with ICM, does not receive FGF4 signalling queues and ceases to proliferate (Gardner et al., 1973, Rossant and Cross, 2001). Mural trophectoderm cells differentiate into the first specialised trophoblast subtype: primary parietal trophoblast giant cells (p-TGCs). Differentiation to p-TGCs is considered to be the default differentiation pathway in the absence of signalling factors (Gardner et al., 1973). After implantation, the polar trophectoderm continues to receive growth factor stimulation from the epiblast, which results in the formation of the stem cell compartments: the extraembryonic ectoderm (ExE) and later the chorionic ectoderm (Tanaka et al., 1998, Uy et al., 2002). As cells move away from FGF4 and TGFβ signalling in the ExE, they begin to differentiate and form the ectoplacental cone

(EPC). The EPC contains a pool of progenitor cells that co-express two antagonistic basic helix-hoop-helix (bHLH) transcription factors: *Ascl2* and *Hand1*. *Hand1* expression promotes trophoblast giant cell (TGC) formation, while *Ascl2* suppresses it (Carney et al., 1993, Scott et al., 2000). Other non-bHLH transcription factors, such as *Snai1*, also inhibits TGC differentiation in the EPC from 7.5 days post coitum (embryonic day 7.5 (E7.5)) (Nakayama et al., 1998).

At gastrulation, at approximately E7.5, ExE cells in closest proximity to the embryo form the chorion. At E8.5, the chorion fuses with the embryonic mesoderm-derived allantois in a process called chorio-allantoic fusion to form an early placenta structure (reviewed in Watson and Cross, 2005). As the placenta continues to develop, foetal blood vessels invade the allantois and the ectoplacental cone is specified as spongiotrophoblast (SpT). At the same time, the chorion differentiates to form the labyrinth layer, which sits directly beneath the SpT layer (Figure 1.1a) (Adamson et al., 2002, Simmons et al., 2008, Simmons and Cross, 2005). By E14.5, the placenta is mature and fully functional. In all, there are three trophoblast layers: an outer p-TGCs layer in direct contact with the maternal decidua, a SpT containing junctional zone and a labyrinth layer in which gas, nutrient and waste exchange occurs (Adamson et al., 2002).

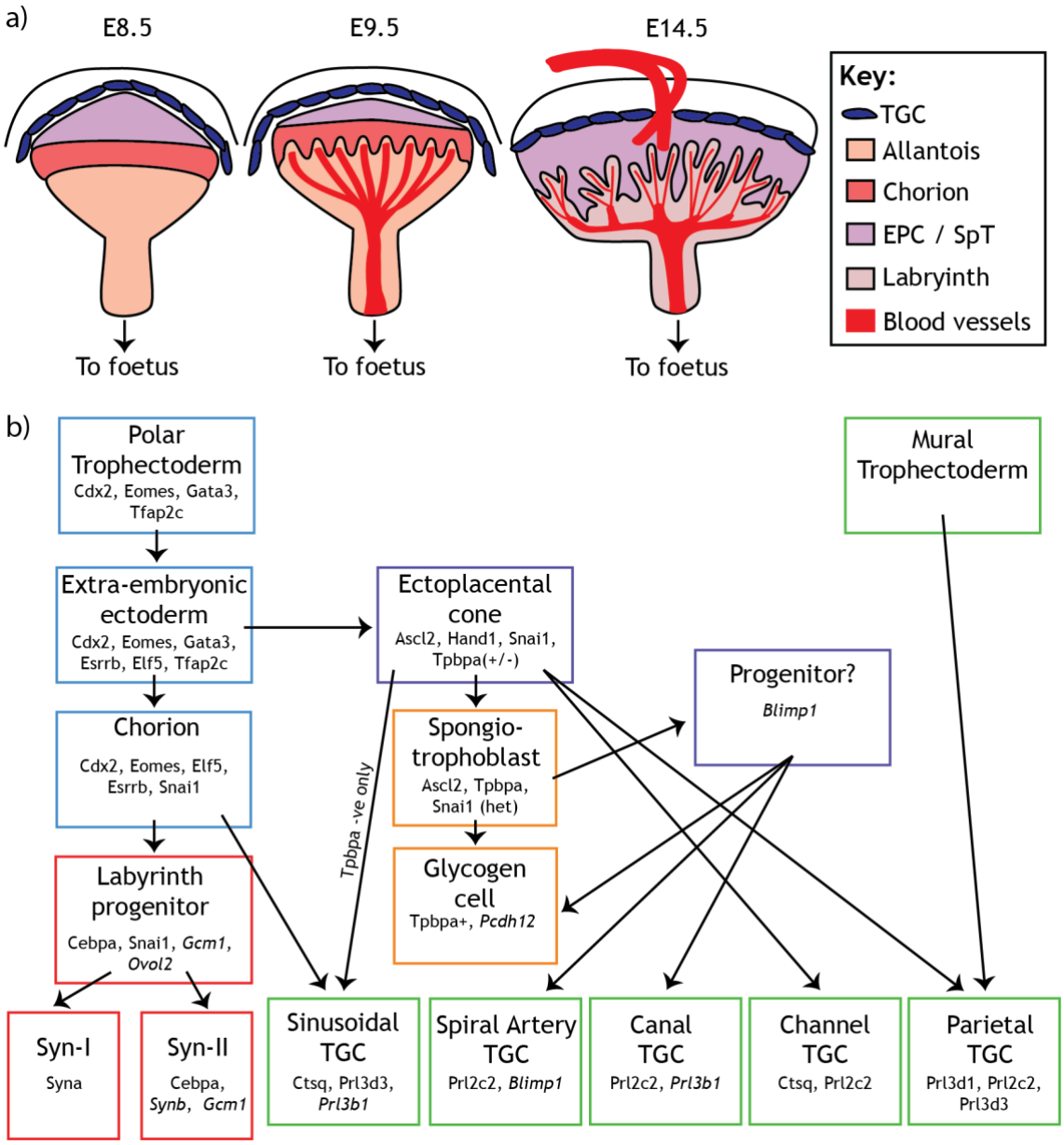



**Figure 1.1: Cell types and structure of mouse placental development.** a) Schematic representing the gross anatomical structure of the mouse placenta immediately after chorio-allantoic fusion (E8.5), an intermediate placenta time point (E9.5) and the mature placenta (E14.5). TGC = trophoblast giant cell, SpT = spongiotrophoblast, EPC = ectoplacental cone. b) An overview of the stem cell (blue), progenitor (purple) and specialised subtypes of the mouse placenta. Non-italicised genes are those that were used in qRT-PCR experiments in this thesis. Red = Syncytiotrophoblast lineage, orange = spongiotrophoblast (SpT), Green = trophoblast giant cell (TGC), het = heterogeneous.

#### *1.2.1.2 Trophoblast giant cells*

There are several specialised TGC subtypes that play unique roles in the placenta (reviewed in (Hemberger, 2008) and (Hu and Cross, 2010)). TGC have different functions, but have some universal characteristics. Firstly, all TGCs express the transcription factor *Hand1*, which is required for the formation of every TGC subtype (Simmons et al., 2007). Secondly, TGCs are large in size (Bevilacqua and Abrahamsohn, 1988). An increase in cell size occurs as TGC undergo successive rounds of endoreduplication (DNA replication without mitoses) resulting in up to 1000 copies of the genome (Geng et al., 2003, Parisi et al., 2003). Lastly, mouse TGCs have a single, large nucleus which contains the endoreduplicated DNA (Simmons et al., 2007). It is unclear why endoreduplication is critical to TGC function. One potential explanation is that polyploidy is advantageous for growth and/or protein synthesis (Cross, 2014, Hemberger, 2008). TGCs secrete a diverse range of proteins, including cell adhesion molecules, proteinases and extracellular matrix; therefore, an increase protein production would be advantageous to cell function (Hemberger, 2008, Hu and Cross, 2010, Rai and Cross, 2014). TGCs are known to secrete prolactins, which differ between TGC subtypes, and are commonly used as a way to identify subgroups of TGCs (Figure 1.1b).

##### 1.2.1.2.1 Parietal trophoblast giant cells

The mural trophectoderm differentiates into primary parietal trophoblast giant cells (p-TGCs). p-TGCs are thought to be important for attachment, implantation and decidua cell differentiation (Gardner et al., 1973, Hu and Cross, 2010). Post-implantation, secondary p-TGCs are generated from the polar trophectoderm (Cross et al., 1994). However, only half of the secondary p-TGCs are derived from *Tpbpa*-expressing progenitor cells in the ectoplacental cone (Simmons et al., 2007). p-TGCs that differentiate from the EPC uniquely express *Prolactin-4a1 (Prl4a1)*, suggesting that this marker could be used to distinguish between progenitor derived p-TGCs and non-progenitor derived pTGCs (Ma and Linzer, 2000). *In vitro*, p-TGCs can also be formed by non-*Tpbpa* expressing precursors in the presence of retinoic acid, but these cells express *Ascl2* during the derivation process, indicating that they still progress through an EPC-like fate (Yan et al., 2001, Simmons et al., 2007).

Additionally, p-TGCs are the only TGC subtype to express *Prolactin-3d1 (Prl3d1)*, which is expressed until E10.5 *in vivo* (Simmons et al., 2007, Rawn et al., 2015).

p-TGCs are components of the parietal yolk sac, which is a transient barrier structure responsible for the initial gas and nutrient exchange between mother and foetus after implantation (Welsh and Enders, 1987, Cross et al., 1994, Nau, 2001). In the established chorio-allantoic placenta, p-TGCs form an outer barrier that is in direct contact with the maternal decidua (Adamson et al., 2002, Hu and Cross, 2010). In all, p-TGCs are thought to be required for a broad range of functions, controlling various aspects of maternal placental physiology, such as vasculature, extracellular matrix and the hormonal changes required to maintain pregnancy (Hemberger, 2008, Hu and Cross, 2010).

#### 1.2.1.2.2 Spiral artery trophoblast giant cells

As their name suggests, spiral artery trophoblast giant cells (SpA-TGCs) line maternal spiral arteries. Present from E8.5 onwards, this TGC subtype is highly invasive and migrates into the maternal decidua. Once migration is complete, the SpA-TGCs secrete factors to regulate arterial remodelling and line maternal spinal arteries (Adamson et al., 2002, Hu and Cross, 2010). As a result, this TGC subtype is critical for facilitating blood flow into the placenta. Lineage tracing studies indicate that almost all SpA-TGCs are derived from *Tpbpa*-positive precursors (Simmons et al., 2007). More recent evidence shows that SpA-TGC precursors express *Blimp1* (Mould et al., 2012).

#### 1.2.1.2.3 Canal trophoblast giant cells

Canal TGCs (c-TGCs) are structurally important for lining the maternal blood canals, (Simmons et al., 2007, Hu and Cross, 2010). The first c-TGCs are specified by E10.5, and they are continually generated to cover the increasing surface area of the maternal blood canals during development (Mould et al., 2012). The developmental origin of c-TGCs is heterogeneous. Whilst half of c-TGCs are derived from *Tpbpa* expressing progenitors in the outer EPC and spongiotrophoblast, the other half are derived from non-*Tpbpa* expressing populations from the inner EPC and chorion

(Simmons et al., 2007). More recent data has shown that c-TGCs are derived from *Blimp1* co-expressing progenitors in the EPC (Mould et al., 2012).

1.2.1.2.4 Sinusoidal trophoblast giant cells

Maternal blood pools in sinusoids within the labyrinth layer of the mature placenta. The sinusoidal spaces are lined by sinusoidal TGCs (s-TGCs) and are the main source of nutrient exchange in the placenta (Rai and Cross, 2014). All s-TGCs come from precursor populations that have never expressed *Tpbpa* (Simmons et al., 2007). Instead, s-TGCs are derived from s-TGC precursors in the chorion that are present from E8.5 onwards (Simmons et al., 2008). s-TGCs are known to secrete hormones, including *Prolactin-2c2 (Prl2c2)* or other proteins such as *Cathepsin Q (Ctsq)*. Given the proximity of s-TGCs to both maternal and foetal circulation, they may play a role in modulating growth factor and hormones (Rai and Cross, 2014, Hu and Cross, 2010).

1.2.1.2.5 Channel trophoblast giant cells

Channel TGCs (Ch-TGCs) form channels that help move maternal blood away from sinusoids in the labyrinth layer after gas and nutrient exchange. The channels attach to p-TGC lined lacunae that are connected to the uterine veins (Rai and Cross, 2014). Thought to derive specifically from differentiation in the EPC, Ch-TGCs are one of only two TGC subtypes to express the marker *Cathepsin Q* (*Ctsq)* and co-express *Prolactin-2c2 (Prl2c2)* (Rai and Cross, 2015).

### *1.2.1.3 Junctional zone cell types*

The junctional zone, derived from the ectoplacental cone, sits between the outer p-TGC layer and inner labyrinth layer (Figure 1.1a) (Adamson et al., 2002, Rai and Cross, 2014). This region contains spongiotrophoblast and glycogen cells.

1.2.1.3.1 Spongiotrophoblast

Spongiotrophoblast (SpT) cells are heterogeneous in appearance over time. Unlike at earlier stages of placental development, E16.5 SpT cells contain extensive endoplasmic reticulum (Coan et al., 2005). Such extensive endoplasmic reticulum may be required to produce the endocrine factors they secrete (Coan et al., 2005).

By E16.5, it is thought that the SpT cells themselves could support the signalling needs until the end of gestation (Coan et al., 2006). Therefore, SpT cells probably play a functional role in modulating maternal and/or foetal physiology (Coan et al., 2005). In addition, SpT may provide structural support for sinusoid development in the underlying labyrinth zone (Simmons and Cross, 2005).

#### 1.2.1.3.2 Glycogen trophoblast

Glycogen trophoblast cell (GTC) progenitors are found in the ectoplacental cone as early as E7.5 (Tesser et al., 2010). Glycogen found in GTC progenitors is metabolised into glucose to serve as an energy source to the migrating GTCs (Tesser et al., 2010). As GTCs mature, the nucleus condenses and glycogen granules form (Coan et al., 2005, Coan et al., 2006). By E12 mature GTCs are found in the maternal decidua having migrated out to this region from the junctional zone (Adamson et al., 2002, Simmons and Cross, 2005). Here, GTCs are thought to be a key source of energy when demand is high at the latter stages of gestation (Coan et al., 2006).

### *1.2.1.4 Labyrinth layer cells*

The labyrinth layer is the site of gas and nutrient exchange in the placenta. Three different chorion derivatives form the barrier between maternal sinusoids and the foetal circulatory system derived from extraembryonic mesoderm (Cross et al., 2006). As discussed above, s-TGCs line the maternal sinusoids. s-TGCs are loosely attached to the adjacent syncytiotrophoblast layer; syncytiotrophoblast-I (Syn-I) (Coan et al., 2005). This allows the Syn-I layer to access maternal blood and nutrients. Syn-I are tightly attached to another syncytiotrophoblast layer, Syn-II, which is in direct contact with the endothelial cells that line the foetal capillaries (Enders, 1965, Hernandez-Verdun, 1974).

#### 1.2.1.4.1 Syncytiotrophoblast

Around the time of chorio-allantoic fusion, at E8.5, the chorion becomes organised into three distinct cell layers that give rise to each of the labyrinth populations: Syn-I, Syn-II and s-TGCs (Simmons et al., 2008, Watson et al., 2011). Whether there is a single progenitor that can differentiate into both Syn subtypes is unclear. Syn-I and

Syn-II express *Syna* and *Synb*, respectively, and *Syna/b* are essential for fusing trophoblast cells to produce multi-nucleated Syn cells (Dupressoir et al., 2005).

### 1.2.2 Differences in placenta development between mouse and human

There have been many recent reviews that highlight the differences between mouse and human placental development (Soncin et al., 2015, Schmidt et al., 2015, Silva and Serakides, 2016). As such, I will only outline a few key differences relevant to this thesis.

The chorio-allantoic placenta develops earlier in humans than in mice (Malassine et al., 2008, Schmidt et al., 2015). Given this, it is not surprising that mouse and human placentas are structured differently. However, it has been suggested that there are structurally and/or functionally analogous cells between mouse and human placentas (see (Soncin et al., 2015)). Cytotrophoblast cells (CTBs) are thought to be the stem cells of the human placenta (Hemberger et al., 2010, Soncin et al., 2018). CDX2 expressing CTBs are particularly abundant in early gestation, near the chorionic plate and co-express the mouse trophoblast stem cell marker ELF5 (Hemberger et al., 2010, Soncin et al., 2018). However, CDX2 expression is absent from CTBs at other places in the placenta (Soncin et al., 2018).

Beyond CDX2 and ELF5, other key mouse trophoblast stem cell markers, including EOMES, are absent (Soncin et al., 2018). This indicates that the transcriptional network used to establish and maintain trophoblast stem cells in humans is different to that used in mice (Soncin et al., 2018). Despite these differences, many of the genes known to be important for TGC formation in mice are conserved in their human equivalent: extravillous trophoblast (EVT) (Rawn and Cross, 2008). Given the ethical implications of studying human placental development *in vivo* and the limited *in vitro* tools available, mouse *in vivo* and *in vitro* models are useful and necessary alternatives.

## 1.3 Derivation and maintenance of *in vitro* trophoblast stem cells

In 1981, stable *in vitro* stem cell lines were derived and propagated from early embryos for the first time (Evans and Kaufman, 1981, Martin, 1981). These so-called mouse embryonic stem cells (mESCs), derived from the ICM, were pluripotent as shown in teratoma carcinomas and chimera embryos (Bradley et al., 1984, Evans et al., 1985, Beddington and Robertson, 1989). Almost two decades later, Janet Rossant and colleagues successfully isolated and maintained self-renewing stem cells from both the extraembryonic ectoderm of E6.5 embryos and trophectoderm of E3.5 embryos (Tanaka et al., 1998). Subsequent work derived similar cells from E8.5 chorionic ectoderm (Uy et al., 2002). These isolated cells, known as trophoblast stem cells (TSCs), are indistinguishable from each other in both appearance and differentiation potential regardless of which stage they are derived from (Tanaka et al., 1998, Uy et al., 2002). When the growth factors that maintain trophoblast stem cell cultures are removed, proliferation declines and cells differentiate to all trophoblast derivatives (Tanaka et al., 1998). When injected into blastocysts, TSCs contribute solely to trophoblast lineage derivatives (Tanaka et al., 1998). Further, TSCs remodel vessels to form of blood-filled lacunas when they are injected under the skin of mice, indicating that TSCs retain their invasiveness and functionality *in vivo*, even outside of the embryo context (Kibschull et al., 2004, Kibschull and Winterhager, 2006).

### 1.3.1 Signalling pathways that maintain trophoblast stem cells

Fibroblast growth factor 4 (FGF4) is a key growth factor required to derive and maintain TSCs (Tanaka et al., 1998). *In vivo*, FGF4 is expressed in the ICM and Epiblast and FGFR2 is expressed in the TE. The importance of FGF4 signalling is confirmed by the observation that *Fgf4*-null and *Fgfr2*-null embryos are not viable (Tanaka et al., 1998, Feldman et al., 1995, Arman et al., 1998). In the trophectoderm, FGF4 activates the RAS/MAPK1 signalling pathway, which is required to maintain stemness but is not required for initial lineage segregation (Feldman et al., 1995, Arman et al., 1998, Saba-El-Leil et al., 2003, Lu et al., 2008). The other components of the original TSC culture medium were less-well defined, containing 20% foetal

bovine serum and factors secreted by mouse embryonic fibroblasts (MEFs) (Tanaka et al., 1998). Subsequent work established that MEFs provide nodal and/or activin; these are both members of the transforming growth factor beta (TGFβ) superfamily and are critical for TSC proliferation and maintenance (Erlebacher et al., 2004, Natale et al., 2009, Saba-El-Leil et al., 2003). In recent years, two independent groups have optimised defined culture conditions in which TSCs can be derived and maintained (Kubaczka et al., 2014, Ohinata and Tsukiyama, 2014). Despite the differences between medium composition, both support differentiation into all trophoblast derivatives both *in vitro* and in chimera experiments *in vivo*.

### 1.3.2 Alternative trophoblast stem cell derivation methods

Work aimed at understanding the reciprocal relationship between *Oct4* and *Cdx2* inadvertently discovered that overexpressing *Cdx2* or repressing *Oct4* in mESCs can convert them into TSC-like cells (Niwa et al., 2000, Niwa et al., 2005). Conversely, overexpression of *Oct4* alone in TSCs can convert them into mESCs, but the efficiency of this is increases by overexpressing *Oct4* in combination with the other core 'Yamanka' factors (Wu et al., 2011, Kuckenberg et al., 2011). The importance of the *Cdx2*-*Oct4* relationship at an early developmental stage has been confirmed by the inability of *Cdx2* overexpression to induce trophoblast gene expression in stem cells derived from the epiblast (Blij et al., 2015). Overall, the ability to 'transdifferentiate' mESCs to TSC-like cells and vice versa demonstrates the importance of the repressive, antagonistic relationship between *Oct4* and *Cdx2* (Niwa et al., 2000, Niwa et al., 2005, Strumpf et al., 2005, Wu et al., 2011, Kuckenberg et al., 2011).

Overexpression of other trophoblast related transcription factors has been reported to convert mESCs to the trophoblast lineage, with varying degrees of success; including *Eomes*, *Elf5*, *Gata3*, *Tfap2c* and *Tead4* (Niwa et al., 2005, Ng et al., 2008, Ralston et al., 2010, Kuckenberg et al., 2010, Nishioka et al., 2009). TSC-like cells can also be generated by overexpressing *Ras* and, therefore, activating MAPK signalling (Lu et al., 2008). The morphological appearance and validation performed for each approach is outlined in Table 1.1.

Recently, two independent groups successfully reprogrammed mouse embryonic fibroblasts (MEFs) into induced-TSCs (iTSCs) by overexpressing a combination of transcription factors: *Gata3*, *Eomes* and *Tfap2c* (GET) with either *Myc* or *Ets2* (Benchetrit et al., 2015, Kubaczka et al., 2015). This process is successful but inefficient, with a success rate of only 0.03% (Benchetrit et al., 2015, Kubaczka et al., 2015). Interestingly, cells undergoing reprogramming do not express pluripotent marker genes, *Oct4, Nanog* or *Sox2*, suggesting that they do not progress through an 'mESC-like' fate. Additionally, Kubaczka et al (2015) found that overexpression of GET + *Ets2* could not convert mESCs into stable proliferating iTSCs (Kubaczka et al., 2015).

TSC-like cells can also be generated by growth factor induced approaches, although these are less commonly used (Hayashi et al., 2010, Schenke-Layland et al., 2007, He et al., 2008). BMP4-treatment of 1-cell to 8-cell stage mouse embryos induces TE markers expression in the nuclei of inner cells pre-compaction and in ICM cells of blastocyst embryos (Home et al., 2012). Similarly, BMP4-treatment of mESCs in defined culture conditions on laminin has also been shown to drive differentiation towards TSC-like cells (Hayashi et al., 2010). Others have shown that Wnt signalling or collagen-V can drive also transdifferentiation of mESCs to TSC-like cells (Schenke-Layland et al., 2007, He et al., 2008). However, growth factor-induced transdifferentiation is heterogeneous, cultures do not resemble TSCs and the validation relies solely on the expression of a relatively small panel of genes (Hayashi et al., 2010, Schenke-Layland et al., 2007, He et al., 2008). As such, growth factor induced TSCs are unlikely to represent *bona fide* TSCs.

| Approach | Cell type | Morphology | Comment | Validation | | Reference |
|---|---|---|---|---|---|---|
| | | | | Epigenetic | Chimeras | |
| Cdx2 overexpression | mESCs | Formed TS-like colonies in TSCMM | Upregulation of Eomes, Esrrb, Ets2 and Hand1 | None | Contribution to placenta only | Niwa et al., 2005 |
| | mESCs | Formed TS-like colonies on feeders | No initial difference in Tead-/- conditions but TSCs cannot be maintained | None | None | Nishioka et al., 2009 |
| | mESCs | Conversion to TS-like morphology in mESC conditions. More prominent when converted to TS | RNA-seq data available | ATAC-seq. Still highly dissimilar to TSCs | None | Rhee et al., 2017 |
| Cdx2 overexpression (CRISPR-mediated activation) | mESCs | WT colonies but some TGC differentiation | Endogenous Cdx2, Tfap2c, Elf5, Tead4, Eomes and Sox2 expressed at similar to in TSCs. Some cell lines retained some Nanog and Oct4 expression | Methylation gained at Oct4, lost at Elf5 | Contribution to placenta only | Wei et al., 2015 |
| Eomes overexpression | mESCs | Large cells in small flat colonies<br>Formed TS-like colonies in TSCMM | Colonies are different to those induced by Cdx2 overexpression | None | None | Niwa et al., 2005 |
| Gata3 overexpression | mESCs | Induced TS-like colonies but this could not be maintained | Also induced non-trophoblast genes | None | None | Ralston et al., 2010 |
| | mESCs | Not specified. Not converted to TS medium. | Induced primitive endoderm genes | None | None | Nishiyima et al., 2009 |
| | mESCs | Conversion to TS-like morphology in mESC conditions. More prominent when converted to TS | RNA-seq data available | ATAC-seq. Still highly dissimilar to TSCs | None | Rhee et al., 2017 |
| Oct4 downregulation (conditional allele) | mESCs | Monolayer colonies of regular epithelial cells in TS medium | Can be maintained for over a month. They differentiate into TGC-like cells after growth factor | None | None | Niwa et al., 2000 |
| Ras overexpression (conditional allele) | mESCs | Formed TS-like colonies on feeders | N/A | None | Contribution to placenta only | Lu et al., 2008 |
| Tead4 overexpression (Tead4-VP16ER) | mESCs | Formed TS-like colonies on feeders indistinguishable from Cdx2 overexpression | Can drive expression of some TE genes in Cdx2-/- but TSCs cannot be maintained | None | None | Nishioka et al., 2009 |
| Tfap2c overexpression | mESCs | TSC colony morphology, enlarged and multiple nuclei | Upregulation of Cdx2, Eomes, Bmp4, Gata3, Tfap2a, Hand1, Prl3d1, and Prl2c2.<br>Transdifferentiate in Cdx2-/- mESCs | None | Contribute to the TE, but embryos never transferred | Kuckenberg et al., 2010 |

**Table 1.1: An overview of the transcription factors used into transdifferentiate mESCs to TSC-like cells and the methods used to validate them.** In validation, 'none' indicates that no such validation was performed.

### 1.3.3 Human trophoblast stem cells: to be or not to be?

For almost two decades, *in vitro* work on human trophoblast has relied on TSC-like cells transdifferentiated from human ESCs using exogenous BMP4 (Xu et al., 2002, Das et al., 2007, Li et al., 2013). When exposed to BMP4, human ESCs decrease expression of pluripotency genes and increase the expression of TE-genes (Amita et al., 2013). However, they cannot be propagated and it has been suggested that the cells generated using this method are extraembryonic mesoderm (Bernardo et al., 2011). Therefore, whether such cultures represent human trophoblast is contentious (Roberts et al., 2014).

Given these problems, others have tried to generate human TSCs from blastocysts or placenta directly (Kunath et al., 2014, Okae et al., 2018). Derivation methods used for mice do not work for human blastocysts, as only one of 60 blastocysts could be cultured for three passages and these cells did not resemble human ESCs nor mouse TSCs (Kunath et al., 2014). Okae et al (2018) recently published a method to derive self-renewing cells from both human blastocysts and post-implantation villous cytotrophoblast (Okae et al., 2018). This approach modulates completely different pathways to those targeted in mouse and does not use BMP4. Whilst these cells do have gene expression profiles similar to primary trophoblast cells, further validation is required to understand what these cells resemble (Okae et al., 2018, Wamaitha and Niakan, 2018).

### 1.3.4 Validating transdifferentiated and reprogrammed trophoblast stem cells

The methods used to validate TSC-like or iTSCs vary between publications. In some cases, transdifferentiation into TSC-like cells was confirmed by the expression of a few trophoblast marker genes (Niwa et al., 2005, Ralston et al., 2010, Nishioka et al., 2009). However, the transcription factors used in these studies are expressed in various developmental lineages and can, therefore, also induce genes associated with these lineages (Ralston et al., 2010, Nishiyama et al., 2009). As such, whether transdifferentiation generates stable, *bona fide* TSCs has been subject to controversy. TSC-like cells derived from *Cdx2* overexpression or *Oct4* knockout are homogenous and phenotypically normal in appearance, but do not resemble *bona fide* TSCs transcriptionally nor epigenetically (Cambuli et al., 2014). Currently, other

than using the transcriptome, there are two described methods of validating transdifferentiation into TSCs: examining contribution to embryos in chimera experiments and/or the epigenome. The validation performed for each transdifferentiation approach is outlined in Table 1.1.

#### *1.3.4.1 Using chimera integration*

To validate *in vitro* TSCs, Tanaka et al (1998) injected cells into blastocysts to determine their contribution to the mouse embryo. This assay revealed that TSCs only contribute to the placenta (Tanaka et al., 1998). Consequently, transdifferentiated TSC-like cells and reprogrammed iTSCs have been validated using chimeras (Table 1.1). While some studies determined their long-term contribution by transferring embryos to pseudo-pregnant females, others were only assessed for the ability to contribute to the trophectoderm *ex vivo* (Table 1.1). TSC-like cells can contribute to the placenta, but this contribution is poorer than for *bona fide* TSCs (Cambuli et al., 2014).

#### *1.3.4.2 Using the epigenome*

TSCs and mESCs have distinct histone modification and DNA methylation profiles (Rugg-Gunn et al., 2010, Senner et al., 2012). The importance of these DNA methylation profiles is apparent from *Dnmt1*-null mESCs and embryos. *Dnmt1*-null mESCs and embryos cannot maintain the hypermethylated status of the *Elf5* locus, resulting in spontaneous TGC-like differentiation *in vitro* and in the embryo proper *in vivo* (Ng et al., 2008). When tested, several loci that are hypomethylated in TSCs, including the key lineage barrier gatekeeping gene *Elf5*, remain methylated in transdifferentiated TSC-like cells (Cambuli et al., 2014). More recently, it has been shown that the chromatin landscape of *bona fide* TSCs and TSC-like cells are highly dissimilar (Rhee et al., 2017).

#### *1.3.4.3 Using the transcriptome*

TSC-like cells transdifferentiated from mESCs are significantly less proliferative than *bona fide* TSCs (Cambuli et al., 2014). TSC-like cells do not express TSC marker genes to the same levels as in *bona fide* TSCs and their expression fluctuates from between passages (Cambuli et al., 2014). Multiple RNA-seq studies have confirmed

that TSC-like cells do not faithfully recapitulate the *bona fide* TSC transcriptome (Adachi et al., 2013, Aiba et al., 2009). It is known that those transcription factors that maintain the TSC gene regulatory network (GRN) are stringently balanced, regulating stem cell maintenance and differentiation (Latos et al., 2015b, Nelson et al., 2017). As such, fluctuations in stem cell marker genes may be symptomatic of an unstable gene regulatory network and reflect incomplete conversion of the cells.

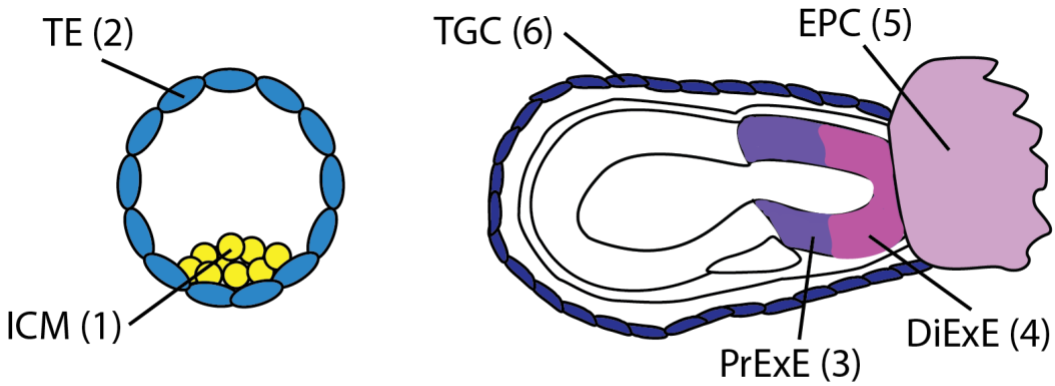
#### *1.3.4.4 Lineage barriers and their incompatibility*

At the time of writing this thesis, only one study has used more than one of these approaches when validating their transdifferentiated mESCs (Wei et al., 2016). Wei et al (2016) used a CRISPR-mediated approach to induce endogenous *Cdx2* expression rather than overexpression. Whilst these cells express TSC marker genes at similar levels as *bona fide* TSCs and show altered methylation, no global transcriptional nor epigenomic analyses were performed. In contrast, MEF reprogramming to iTSCs using GET, with or without *Ets2* or *Myc*, faithfully reprograms DNA methylation marks and histone modifications globally (Kubaczka et al., 2015, Benchetrit et al., 2015). As such, iTSCs are nearly indistinguishable from *bona fide* TSCs in their transcriptome, morphology and DNA methylation status (Benchetrit et al., 2015, Kubaczka et al., 2015). In all, this confirms that TSCs can be generated using alternative approaches to embryonic-outgrowths.

Kubaczka et al (2015) offered two potential explanations for this incompatibility in generating TSCs from mESCs by overexpression: 1) The DNA methylation status is more similar between TSCs and MEFs, resulting in easier conversion between these fates; 2) To prevent aberrant differentiation between early developmental lineages in the blastocyst, there is tighter regulation of the epigenetic barrier (Kubaczka et al., 2015). In contrast, *Oct4*, alone or in combination with other factors, can overcome this epigenetic barrier and transdifferentiate TSCs into epigenetically and transcriptionally indistinguishable mESCs (Kuckenberg et al., 2011, Wu et al., 2011). Thus, it is possible that we have not yet determined the correct conditions to overcome the lineage barrier between mESCs and TSCS.

## 1.4 Deciphering the trophoblast stem cell gene regulatory network

When and where the transcription factors that establish and maintain TSCs are expressed in early development varies greatly. Some TSC markers are expressed in both the TE or ICM, highlighting the importance of context for their functionality. Here, I provide an overview of the expression of core TSC marker genes in early development (Figure 1.2), their placental knockout phenotype and their known functions in trophoblast stem cells.




| | 1 | | 2 | | 3 | | 4 | | 5 | | 6 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | T | P | T | P | T | P | T | P | T | P | T | P |
| Cdx2 | | | | | | | | | | | | |
| Elf5 | | | | | | | | | | | | |
| Eomes | | | | | | | | | | | | |
| Esrrb | | | | | | | | | | | | |
| Ets2 | | | | | | | | | | | | |
| Gata3 | | | | | | | | | | | | |
| Sox2 | | | | | | | | | | | | |
| Tead4 | | | | | | | | | | | | |
| Tfap2c | | | | | | | | | | | | |

**Figure 1.2: A summary of trophoblast stem cell marker gene expression in early placental development.** The top left schematic represents an E3.5 blastocyst and top right is a E6.5 embryo. ICM = inner cell mass, TE = trophectoderm, PrExE = proximal extraembryonic ectoderm, DiExE = distal extraembryonic ectoderm, EPC = ecoplacental cone, TGC = trophoblast giant cell, T = Transcript, P = protein.

### 1.4.1 Cdx2

*Cdx2* is first expressed at the 8- to 16-cell stage in mice. CDX2 protein is present in the outer cells of the 16-cell embryo and subsequently in the trophectoderm of the blastocyst (Beck et al., 1995, Guo et al., 2010, Strumpf et al., 2005). Peri-implantation, at approximately E4.5, CDX2 expression is downregulated in the mural trophectoderm (Strumpf et al., 2005, Frias-Aldeguer et al., 2019). In the polar trophectoderm and its stem cell derivatives, the extraembryonic ectoderm and subsequently the chorionic ectoderm, CDX2 expression is maintained (Beck et al., 1995). Later, at E7.5-8.5, *Cdx2* expression is present in ectoplacental glycogen cells and at E9.5 in the spongiotrophoblast (Beck et al., 1995, van Nes et al., 2006). Although *Cdx2*-null embryos can form blastocysts, they cannot maintain epithelial integrity. As a result, their blastocoel cavity collapses and embryos die before implantation (Strumpf et al., 2005). CDX2 antagonises *Nanog* and *Oct4* expression by binding and repressing their promoters (Niwa et al., 2005, Chen et al., 2009b). As such, both *Nanog* and *Oct4* are expressed in all cells of *Cdx2*-null embryos (Strumpf et al., 2005).

*Cdx2* expression is driven by different enhancers at different developmental stages (Rayon et al., 2016). As outlined in Chapter 1.1.1, *Cdx2* expression is activated by TEAD4 downstream of Hippo signalling (Nishioka et al., 2009). Recently, it has been shown that Notch signalling, which is specifically active in outer cells of the compacting blastocyst, also induces *Cdx2* expression (Rayon et al., 2014). Hippo (via YAP) and Notch (via RBPJ) signalling pathways converge on the TE-specific enhancer (TEE) which is bound by TEAD4 and co-activate *Cdx2* expression (Rayon et al., 2014, Rayon et al., 2016). Activation of the TEE enhancer is also dependent on the chromatin factor SBNO1, which is thought to be required to recruit the histone chaperone FACT (Watanabe et al., 2017). After implantation and in *in vitro* TSC cultures, the TEE enhancer is inactive. Instead, *Cdx2* expression is controlled using other enhancers that may also be active in the blastocyst (Rayon et al., 2016). Although it is not directly involved in activating the TEE, the TSC transcription factor *Tfap2c* is also important for Hippo signalling, suppressing its activity via the cell polarity protein PARD6B (Cao et al., 2015). Independent of this, TFAP2C is also responsible for regulating *Cdx2* activity from an enhancer in intron 1 of *Cdx2* (Cao et

al., 2015). Additionally, it has been suggested that *Cdx2* can also autoregulate its own expression (Niwa et al., 2005).

Although *Cdx2* function is now known to be downstream of initial TE commitment, it is still essential for TSC self-renewal (Ralston and Rossant, 2008, Wu et al., 2010). Trophoblast outgrowths cannot be generated from *Cdx2*-null embryos, which never express trophectoderm markers nor markers of trophoblast differentiation (Strumpf et al., 2005). TSC-like cells cannot be maintained without *Cdx2*, instead upregulating differentiation markers and subsequently terminally differentiating (Niwa et al., 2005, Kuckenberg et al., 2010). Similarly, *Cdx2* knockdown in TSCs downregulates TSC marker genes and upregulates the TGC and Syn-I markers *Prl3d1* and *Syna*, respectively (Table 1.2) (Latos et al., 2015b). This is unsurprising as *Cdx2* is downregulated during TSC differentiation *in vitro* and *in vivo* (Beck et al., 1995, Tanaka et al., 1998).

The role of CDX2 in the trophoblast gene regulatory network is unclear. CDX2 is known to directly interact with other TSC marker genes, including TEAD4 and EOMES (Latos et al., 2015a). As such, it has been speculated that these might represent CDX2 co-factors within the trophoblast context (Huang et al., 2017). Others have shown that CDX2 binds endogenous retroviruses, known to be core enhancers that drive placental evolution and divergence between species, in conjunction with EOMES and ELF5 (Chuong et al., 2013). In addition to these TSC specific roles, it has been suggested that CDX2 can drive the expression of *Bmp4* from the ExE that is required to maintain the epiblast (Murohashi et al., 2010).

### 1.4.2 Elf5

ETS transcription factor family member *Elf5* is first expressed in the trophectoderm of pre-implantation embryos, but ELF5 protein is not detected at these stages (Donnison et al., 2005, Ng et al., 2008). Post-implantation, *Elf5* is expressed in the proximal and distal extraembryonic ectoderm (ExE) and chorionic ectoderm (Donnison et al., 2005, Latos et al., 2015b, Ng et al., 2008). *Elf5* expression levels are critical for early embryonic development. Whilst *Cdx2* is expressed normally in pre-implantation *Elf5*-null embryos, post-implantation embryos develop smaller ExE

compartments that do not express *Cdx2*. As a result, *Elf5*-null embryos die before E10.5 due to placental defects (Donnison et al., 2005). *Elf5* overexpression *in vivo* reduces the number of *Tpbpa*-positive spongiotrophoblast precursors and induces a vast overabundance of TGCs, resulting in embryo reabsorption by E10.5 (Latos et al., 2015b). This suggests that *Elf5* expression must be tightly regulated.

In agreement with the *in vivo* phenotype, trophoblast outgrowths from *Elf5*-null embryos only contain cells resembling TGCs (Donnison et al., 2005). Similarly, knockdown of *Elf5* in TSCs downregulates stem cell markers such as *Eomes* and *Cdx2* and upregulates TGC and Syn markers (Latos et al., 2015b). However, neither *Eomes* nor *Cdx2* are direct targets of *Elf5*, with only *Sox2* expression downregulated 24-hours post *Elf5* knockdown (Table 1.2) (Pearton et al., 2014). *Elf5* overexpression in mESCs drives them to differentiate towards the trophoblast lineage, but these TSC-like cells do not self-renew (Ng et al., 2008). Likewise, TSCs overexpressing *Elf5* can be maintained, but these cultures are unstable and show an increased propensity to differentiate into TGCs (Latos et al., 2015b). Therefore, the requirement for exact *Elf5* expression levels *in vivo* are also replicated in *in vitro* cultures.

One potential explanation for these phenotypes is that ELF5 interacts with different proteins depending on its expression level. In TSCs, ELF5 preferentially interacts with EOMES, but also interacts with TFAP2C. All three transcription factors co-bind and activate key TSC marker genes, including themselves and *Cdx2* (Latos et al., 2015b, Ng et al., 2008). During *in vitro* differentiation, *Elf5* is briefly upregulated before it is downregulated (Donnison et al., 2015, Latos et al., 2015b). In differentiation, ELF5 preferentially interacts with TFAP2C and binds around genes associated with trophoblast differentiation (Latos et al., 2015a). In combination, this explains why TSCs cannot be maintained with too much or too little *Elf5* expressed.

### 1.4.3 Eomesodermin (Eomes)

A T-Box transcription factor, *Eomes* is expressed in the trophectoderm and persists in the ExE of post-implantation embryos (Ciruna and Rossant, 1999, Hancock et al., 1999, Russ et al., 2000, Arnold et al., 2009). Although *Eomes*-null embryos implant, they remain blastocyst-like in appearance, suggesting that trophoblast differentiation

is compromised (Russ et al., 2000, Strumpf et al., 2005). Unsurprisingly, TSC cultures cannot be generated *ex vivo* from *Eomes*-null embryos as they remain as blastocyst-like structures and do not generate trophoblast outgrowths (Russ et al., 2000, Strumpf et al., 2005). Whilst *Cdx2* is not required to initiate *Eomes* expression, it cannot be upregulated in the absence of *Cdx2*, indicating that *Eomes* is downstream of *Cdx2* in TE establishment (Strumpf et al., 2005).

*Eomes* is a core member of the TSC GRN, forming interactions with a number of other transcription factors. *Eomes* knockdown in TSCs results in differentiation towards the trophoblast giant cell (TGC) and syncytiotrophoblast lineages (Kidder and Palmer, 2010, Latos et al., 2015b). *Sox2* is downregulated two days after *Eomes* is knocked-down and is likely a direct target of its expression. Downregulation of *Cdx2* and *Elf5* only occurs six days after *Eomes* expression is knocked-down, showing that their dysregulation is an indirect effect of *Eomes* loss (Table 1.2) (Kidder and Palmer, 2010, Latos et al., 2015b). As discussed above, stoichiometric ratios between *Eomes*, *Elf5* and *Tfap2c* regulate the maintenance or differentiation of TSCs (Latos et al., 2015b). Together, these transcription factors bind a number of TSC marker gene promoters including the *Eomes* promoter, suggesting that EOMES autoregulates itself (Kidder and Palmer, 2010, Latos et al., 2015b).

### 1.4.4 Esrrb

*Esrrb* is highly expressed in all cells of the embryo up to the 16-cell stage. Subsequently, *Esrrb* expression is restricted to the ICM of the mid-blastocyst and epiblast of the late-blastocyst but is absent from the TE (Guo et al., 2010, Adachi et al., 2013, Boroviak et al., 2014). As in the ICM, *Esrrb* is expressed in mESCs, induced by the GSK3 (Glycogen Synthase Kinase-3) signalling pathway, and is both necessary and sufficient to stabilise a naïve pluripotent state (Martello et al., 2012). In this context, ESRRB, facilitated by OCT4, binds the promoter regions of other core pluripotency genes including *Nanog* (van den Berg et al., 2008, Festuccia et al., 2012). *Esrrb* knockdown or knockout in *in vitro* mESCs results in a reduction in mESC self-renewing capacity (Ivanova et al., 2006, Festuccia et al., 2012). Despite this essential role in maintaining pluripotency *in vitro*, *Esrrb*-null embryos survive *in utero* until approximately E10.5 (Luo et al., 1997).

Post-implantation, *Esrrb* is expressed in the proximal extraembryonic ectoderm from E5.5 and later in the chorionic ectoderm (Luo et al., 1997, Adachi et al., 2013). *Esrrb*-null embryos show severe placental abnormalities, including a gross expansion of secondary TGCs at the expense of other trophoblast lineages and a failure of chorio-allantoic fusion (Luo et al., 1997). *Esrrb* is expressed in TSCs and is essential to TSC maintenance (Tanaka et al., 1998). Addition of the oestrogen-related receptor antagonist Diethylstilbestrol (DES) or *Esrrb* knockdown in TSC cultures drives differentiation specifically towards the TGC lineage as observed *in vivo* (Tremblay et al., 2001, Latos et al., 2015a, Gao et al., 2019). Short-term interference of ESRRB activity downregulates several key TSC marker genes, including *Eomes*, *Elf5* and *Sox2*, suggesting that *Esrrb* directly regulates them (Table 1.2) (Latos et al., 2015a). In keeping with this, ESRRB binds the promoter-TSS of key TSC maintenance genes and drives their expression (Adachi et al., 2013, Latos et al., 2015a). However, short-term *Esrrb* perturbation minimally affects *Cdx2* expression suggesting it is not a direct target of *Esrrb* in TSC maintenance (Latos et al., 2015a).

*Esrrb* expression is required for transdifferentiation of MEFs into induced TSCs (iTSCS), and can replace *Eomes* in the cocktail of factors required for this process (Gao et al., 2019). In TSCs, *Esrrb* expression is downstream of FGF/ERK signalling and, when overexpressed alone (Gao et al., 2019), or with *Sox2* (Adachi et al., 2013), it can substitute for exogenous FGF4. This suggests that *Esrrb* is an effector of FGF/ERK signalling, controlling the expression of other key TSC marker genes (Latos et al., 2015a). Unlike in mESCs, ESRRB interacts with *Kdm1a* in TSCs. This is a histone demethylase that regulates the onset of trophoblast differentiation and co-binds many ESRRB binding sites (Zhu et al., 2014, Latos et al., 2015a).

### 1.4.5 Ets2

The ETS transcription factor family member *Ets2* is expressed in the ExE and EPC. *Ets2* expression increases in both of these lineages from E5 to E6.75, and then is downregulated from E7.75 (Georgiades and Rossant, 2006, Yamamoto et al., 1998). Despite its short temporal expression, loss of *Ets2* causes embryos to die at E8.5-9.5. Death of *Ets2*-null embryos occurs as a result of a defective placenta because

tetraploid aggregations can rescue this phenotype (Georgiades and Rossant, 2006, Yamamoto et al., 1998). *Ets2*-null embryos fall into two categories depending on phenotype: Type-I and Type-II. Type-I embryos contain no ExE and a small region resembling the EPC. In contrast, Type-II embryos possess an EPC and an ExE-like domain, expressing associated marker genes *Cdx2*, *Eomes*, *Esrrb* and *Sox2* for a short period of time (Georgiades and Rossant, 2006, Polydorou and Georgiades, 2013).

In line with *in vivo* data, *Ets2* knockdown or knockout in TSCs downregulates a small number of genes including *Cdx2*, *Eomes* and *Esrrb* (Odiatis and Georgiades, 2010, Wen et al., 2007). This observed downregulation of *Cdx2* is likely driven by the loss of ETS2 binding at a functional binding site in the distal *Cdx2* promoter region (Wen et al., 2007). *Ets2* expression in TSCs is required to promote self-renewal, maintain TSC fate and allow for differentiation towards the TGC and spongiotrophoblast lineages (Odiatis and Georgiades, 2010, Wen et al., 2007). As *Cdx2* overexpression cannot rescue the self-renewal phenotype, *Ets2* has other roles in TSCs than simply maintaining *Cdx2* expression (Wen et al., 2007). In *Ets2* knockout TSCs, *Hand1* is not upregulated even in the absence of growth factors and TSCs show increased propensity to differentiate towards labyrinth trophoblast (Odiatis and Georgiades, 2010, Wen et al., 2007). In contrast, short-term knockdown of *Ets2* expression does not affect *Elf5* nor *Fgfr2* expression (Wen et al., 2007). Beyond these TSC-specific roles, *Ets2* expression in the ExE is also essential for the production of *Bmp4*, which is required for gastrulation (Polydorou and Georgiades, 2013).

### 1.4.6 Gata3

*Gata3* is first expressed in the 4-cell stage embryo and GATA3 protein in the 8-cell stage embryo. Subsequently, *Gata3* becomes restricted to the outer cells and then trophectoderm of the blastocyst (Home et al., 2009, Ralston et al., 2010). *Gata3* expression is downstream of TEAD4 and the Hippo signalling pathway and is activated in parallel with *Cdx2* (Ralston et al., 2010). Similar to *Cdx2*, *Gata3* is also repressed in the ICM by OCT4 (Ralston et al., 2010). After implantation, GATA3 is expressed in the extraembryonic ectoderm, ectoplacental cone and the trophoblast giant cells that surrounds the embryonic cavity (George et al., 1994, Ralston et al.,

2010, Ma et al., 1997). There is contradictory evidence on the importance of *Gata3* to blastocyst formation. *Gata3*-null embryos develop normally until E9.5 but die at approximately E12 as a result of haematopoietic defects (Pandolfi et al., 1995). However, approximately 65% of embryos in which *Gata3* was knocked-down at the 2-cell stage did not mature to the blastocyst stage (Home et al., 2009). GATA3 directly represses *Gata2* expression in TSCs (Ray et al., 2009). As *Gata2/3* double knockout embryos undergo cavitation but do not implant (Home et al., 2017), perhaps *Gata2* compensates for *Gata3* under knockout, but not knockdown, conditions.

At later stages of development, both *Gata3* and *Gata2* are expressed in TGCs and drive associated *Prolactin* gene expression (Ng et al., 1994). Accordingly, although *Gata3*-null placentas do not show an obvious placenta phenotype at E10.5, both *Prl3d1* and *Prl2c2* are downregulated (Ma et al., 1997).

*Gata3* is expressed in, and essential to, TSCs (Home et al., 2009). In TSCs, GATA3 binds a small number of active promoters (Kidder and Palmer, 2010), as well as a conserved GATA motif in intron 1 of *Cdx2* from which it regulates *Cdx2* expression in both TSCs and TE (Home et al., 2009). *Gata3* is upregulated during differentiation and is important for TGC formation, as shown by the increase in TGCs when *Gata3* is overexpressed (Table 1.2) (Ralston et al., 2010). Therefore, the level of *Gata3* determines whether stem cells are maintained or differentiate (Ralston et al., 2010). In agreement with the tight requirement of *Gata3* expression, self-renewing TSC-like cells cannot be generated by overexpressing *Gata3* in mESCs (Ralston et al., 2010).

### 1.4.7 Sox2

*Sox2* is initially maternally expressed, but rapidly declines with blastocyst cell divisions (Avilion et al., 2003, Guo et al., 2010). Published data on SOX2 dynamics in early mouse embryonic development is contradictory. Whilst some papers suggest that SOX2 expression is constant and universal, others suggest SOX2 is uniquely upregulated in ICM precursors at the 16-cell stage and patterned in parallel with *Cdx2* (Avilion et al., 2003, Keramari et al., 2010, Guo et al., 2010, Wicklow et al., 2014). In blastocysts, SOX2 expression remains nuclear and restricted to the ICM

(Wicklow et al., 2014). Post-implantation, *Sox2* is expressed in a narrow band in the ExE at the embryonic/extraembryonic boundary and later in the extraembryonic ectoderm proportion of the chorion (Avilion et al., 2003, Adachi et al., 2013). *Sox2*-null embryos die shortly after implantation without an ExE or epiblast. Injection of wild-type mESCs into *Sox2*-null blastocysts does not rescue this phenotype, confirming that *Sox2* expression is critical to the trophoblast lineage (Avilion et al., 2003).

*Sox2*-null blastocysts can form trophoblast outgrowths, but these cannot be maintained and differentiate into trophoblast derivatives (Avilion et al., 2003). In established TSCs, *Sox2* is a primary target of FGF signalling (Latos et al., 2015a, Adachi et al., 2013). Combined overexpression of *Sox2* and *Esrrb* can maintain TSC self-renewal and trophoblast identity in the absence of FGF4 signalling (Adachi et al., 2013). *Sox2* expression is also rapidly downregulated after *Elf5* knockdown, suggesting *Sox2* is a direct target of *Elf5* (Table 1.2) (Pearton et al., 2014).

In all, this suggests that *Sox2* initially represses trophectoderm establishment but is later essential for maintaining the established trophoblast stem cell fate. This change in *Sox2* behaviour has been confirmed by experiments in mESCs. When *Sox2* is knocked-down in mESCs, trophoblast marker genes such as *Hand1*, *Eomes*, *Ascl2* and *Prl3d1* are upregulated. However, this is heterogeneous as *Sox2* knockdown also upregulates markers of other embryonic and extraembryonic lineages (Ivanova et al., 2006, Masui et al., 2007). In contrast, if *Sox2* expression is blocked during mESCs conversion to TSC-like cells by *Cdx2* overexpression, few TSC markers genes are expressed. Instead, genes associated with trophoblast differentiation are upregulated (Adachi et al., 2013). To understand how *Sox2* plays opposite roles in these contexts, Adachi et al (2013) generated SOX2 ChIP-seq libraries during the conversion of mESCs into TSC-like cells by *Oct4*-repression (Adachi et al., 2013). These libraries suggest that SOX2 binding is different between mESCs and TSCs. Whilst SOX2 binding sites are enriched for OCT-SOX motifs in mESCs, those in TSC-like cells are enriched for canonical SOX and AP2 motifs. As such, TFAP2C appears to be a core co-factor for SOX2 binding in the trophoblast fate (Adachi et al., 2013).

### 1.4.8 Tead4

*Tead4*, a TEAD family member, is first upregulated at the 4-cell stage and persists in all cells of the pre-implantation embryo. TEAD4 expression is detected in both the TE and ICM of blastocysts (Nishioka et al., 2008, Yagi et al., 2007). There is conflicting data for the subcellular localisation of TEAD4 in the ICM of mice (Nishioka et al., 2008, Home et al., 2012), but it is cytoplasmic in other species such as human, cattle, rats and rhesus monkeys (Home et al., 2012). Post-implantation, *Tead4* is expressed in the extraembryonic ectoderm, ectoplacental cone, chorion and trophoblast giant cells (Yagi et al., 2007). As discussed in Chapter 1.1.1, TEAD4 and the Hippo signalling pathway regulate *Cdx2* and *Gata3* expression during trophectoderm specification (Nishioka et al., 2009, Ralston et al., 2010). *Tead4*-null embryos are morphologically indistinguishable from wild-type embryos until compaction, polarising and forming normal adherens junction. However, *Tead4*-null embryos do not form a blastocoel cavity and express minimal, if any, trophectoderm marker genes (Nishioka et al., 2008, Yagi et al., 2007). Instead, all cells in *Tead4*-null embryos express pluripotency markers OCT4 and NANOG, suggesting that they have all been specified as ICM (Nishioka et al., 2008).

As expected given the *in vivo* phenotype, *Tead4*-null embryos cannot form trophoblast outgrowths (Nishioka et al., 2008, Yagi et al., 2007). *Tead4* is expressed in both TSCs, where it autoregulates its own expression, and differentiated subtypes such as TGCs (Home et al., 2012). In TSCs, *Tead4* knockdown downregulates multiple trophoblast stem cell marker genes, including *Cdx2*, *Gata3*, *Elf5* and *Eomes* (Home et al., 2012). TEAD4 binding sites are found near all of these downregulated genes in both TSCs and blastocysts (Home et al., 2012). TEAD4 binding sites in TSCs are also enriched for the CDX2, EOMES, TFAP2C and SOX2 consensus binding motifs, suggesting that they may be TEAD4 co-factors (Home et al., 2012). Beyond its importance in regulating TSC marker genes, recent reports suggest that TEAD4 is critical for binding and regulating mitochondrial DNA transcription; regulating energy homeostasis in both TSCs and pre-implantation mouse embryos (Kumar et al., 2018, Kaneko and DePamphilis, 2013).

### 1.4.9 Tfap2c

*Tfap2c* is expressed in all cells of the preimplantation embryo but is restricted to the ExE, EPC and TGCs after implantation (Auman et al., 2002, Werling and Schorle, 2002). Pre-implantation, *Tfap2c* is critical for blastocyst formation, regulating cell polarity, proliferation, fluid accumulation and tight junction gene expression and organisation (Choi et al., 2012). Despite this, *Tfap2c*-null embryos undergo compaction and cavitation normally and are able to implant. This lack of phenotype is likely due to functional redundancy with *Tfap2a* given that double *Tfap2a*/*c* mutants die earlier than *Tfap2c* mutants (Winger et al., 2006). After implantation, *Tfap2c*-null embryo placentas have fewer TGCs and no labyrinth layer, resulting in severe growth defects and death by E9.5 (Auman et al., 2002, Werling and Schorle, 2002). The embryonic phenotype caused by *Tfap2c* knockout is rescued by tetraploid aggregation, indicating that embryonic defects occur as a result of the observed placental defects (Auman et al., 2002). *Tfap2c*-null placental defects are likely caused by their ExE, which is either absent or does not express the key TSC markers *Cdx2* and *Eomes* (Auman et al., 2002). *Tfap2c* expression levels are critical to trophoblast development as 60% of *Tfap2c*-heterozygotes are lost after E14.5. Heterozygote embryos have altered placental morphology, decreased expression of trophoblast stem cell and spongiotrophoblast markers and a shifted fate towards s-TGCs (Kaiser et al., 2015).

*In vitro, Tfap2c*-null embryos have the capacity to form trophoblast outgrowths, but these are smaller, have abnormal cell morphology and contain fewer TGCs than those derived from wild-type embryos (Auman et al., 2002, Kuckenberg et al., 2010, Werling and Schorle, 2002). In addition, *Tfap2c*-null or heterozygous trophoblast outgrowths are lost within 10 passages (Kuckenberg et al., 2010). Similarly, *Tfap2c* knockdown in TSCs decreases trophoblast stem cell marker gene expression, changing TSC colony morphology and increasing differentiation (Kidder and Palmer, 2010, Adachi et al., 2013). This is likely caused by changes to the stoichiometric ratios of *Elf5*, *Eomes* and *Tfap2c* that are required to maintain TSCs. Other than maintaining TSCs, *Tfap2c* is also upregulated during trophoblast differentiation (Latos et al., 2015b). Under these conditions, TFAP2C and ELF5 drive a global increased chromatin accessibility, binding around genes involved in differentiation

(Latos et al., 2015b, Nelson et al., 2017). Thus, maintaining *Tfap2c* expression is critical for both TSC maintenance and differentiation.

TFAP2C binds to the promoters of a number of marker genes in TSCs, including *Tfap2c*, *Cdx2*, *Elf5*, *Eomes*, and *Gata3* (Kidder and Palmer, 2010, Kuckenberg et al., 2010). Whilst TFAP2C binding can activate its own and *Cdx2* expression, it has little effect on *Eomes* and *Elf5* expression (Kidder and Palmer, 2010, Kuckenberg et al., 2010). In agreement with this, *Tfap2c* overexpression in mESCs transdifferentiates cells into TSC-like cells and induces *Cdx2* expression and *Tfap2c* knockdown downregulates *Cdx2* (Table 1.2) (Kuckenberg et al., 2010). Furthermore, *Tfap2c* overexpression can transdifferentiate *Cdx2*-null mESCs into a trophoblast-like fate, but *Eomes* and *Elf5* expression is low and these cells cannot be maintained (Kuckenberg et al., 2010). In part, this may be due to a downregulation of *Oct4* during this process and re-distribution of *Sox2* to its TSC-binding targets (Adachi et al., 2013).

### 1.4.10 A summary of transcription factor interplay in trophoblast stem cell maintenance

To generate a network map that displays the interplay between different TSC marker genes accurately would be difficult. Instead, I have made a table that summarises the reported effects of upregulating or downregulating each of these marker genes in TSCs (Table 1.2). Many of these studies only provide information on the effects of TSC marker perturbation after several days. As such, it is difficult to know if observed effects are direct or indirect. Nonetheless, changes to *Cdx2* expression are reported when the expression of every transcription factor listed in Table 1.2 is perturbed. Therefore, *Cdx2* expression is highly dependent on the expression of other TSC maker genes, although this is probably an indirect effect in some cases. Furthermore, changes to almost any component of the core TSC gene regulatory network dysregulates it and drives TSCs to upregulate markers of differentiation (Table 1.2).

| TF | Approach | Time | Affected genes | Morphology | Comment | Reference |
|---|---|---|---|---|---|---|
| **Cdx2** | shRNA knockdown | 5 days | ↓ Elf5, Eomes, Esrrb, Fgfr2<br>↑ Prl3d1, Syna | No comment | N/A | Latos et al., 2015 |
| **Elf5** | shRNA knockdown | 6 weeks ? | ↓ Cdx2, Eomes<br>↑ TGC markers, Syna, Synb | No comment | RNA-seq performed (no gene list provided) | Latos et al., 2015 |
| | shRNA knockdown | 24 hours and 48 hours | ↓ Sox2<br>↑ Hand1 (24 hours only) | Increased differentiation after 6 days | N/A | Pearton et al., 2014 |
| | Overexpression | Not specified, but long-term | ↑ Cdx2 and TGC markers inconsistent between figures<br>↑ Eomes, Tfap2c, Hand1, Ascl2<br>↓ Gcm1, Synb | Increased differentiation | RNA-seq performed (no gene list provided) | Latos et al., 2015 |
| **Eomes** | shRNA knockdown | 6 weeks ? | ↓ Elf5<br>↑ TGC markers, Syna, Synb | No comment | RNA-seq performed (no gene list provided)<br>Similar to differentiation | Latos et al., 2015 |
| | siRNA knockdown (day D0, D2 and D4) | Collected at days D2, D4 and D6 | D2: ↓ Sox2<br>D6: ↓ Cdx2, Elf5, Sox2, Tfap2c | Loss of TS morphology | Knockdown efficiency >50% by D6 | Kidder and Palmer., 2010 |
| | Overexpression | 6 weeks? | Most similar to TSCs.<br>Slight downregulation of Cdx2 and Spry4 | Normal TSC morphology | RNA-seq performed (no gene list provided) | Latos et al., 2015 |
| **Esrrb** | Chemical inhibitor: Diethylstilbestrol | 4 days | ↓ Eomes<br>↑ Prl3d1 | Differentiation into TGCs | N/A | Tremblay et al., 2001 |
| | Chemical inhibitor: Diethylstilbestrol | 24 hours | ↓ Elf5, Eomes, Sox2<br>↑ Prl3d1and other Prl genes and Syna | No comment | RNA-seq performed at 24 hours and 4 days post-treatment. Gene lists provided | Latos et al., 2015 |
| | shRNA knockdown | 5 days | ↓ Cdx2, Elf5, Eomes, Sox2<br>↑ Prl3d1and other Prl genes and Syna | Differentiation into TGCs | RNA-seq performed.<br>Gene lists provided | |
| | shRNA knockdown | Stable transfection | ↓ Cdx2, Elf5, Eomes<br>↑ Prl3d1, Prl3b1 | Visibly differentiating | Efficient knockdown | Gao et al., 2019 |
| **Ets2** | knockout | 1 day and 3 days | ↓ Cdx2, Esrrb, Elf5 | No comment | Colony formation affected.<br>RNA-seq performed (no gene list provided) | Wen et al., 2007 |
| | shRNA knockdown | 8 days post-transduction | ↓ Cdx2, Esrrb | Loss of TS morphology and TGC formation | ~75% reduction in Ets2 expression | Odiatis et al., 2010 |
| **Gata3** | shRNA knockdown | 5 days total (48 hours + 3 days puromycin) | ↑ Tead4<br>↓ Cdx2 | No comment | N/A | Home et al., 2009 |
| | shRNA knockdown | Stable transfection | ↓ Cdx2, Elf5, Eomes | Visibly differentiating | Efficient knockdown | Gao et al., 2019 |
| | Overexpression (inducible Gata3-ER) | 5 days | ↑ Prl2c2<br>↓ Cdx2 | Differentiation into TGCs | Lower doses do not induce differentiation | Ralston et al., 2010 |
| **Sox2** | shRNA knockdown | Stable transfection | ↓ Cdx2, Elf5, Eomes | Visibly differentiating | Efficient knockdown | Gao et al., 2019 |
| **Tead4** | shRNA knockdown | 3 days | ↓ Cdx2, Gata3, Bmp4, Elf5, Eomes, Fgfr2, JunB | No comment | N/A | Home et al., 2012 |
| **Tfap2c** | siRNA knockdown (day D0, D2 and D4) | Collected at days D2, D4 and D6 | D2: ↓ Cdx2 lost by D4<br>D6: ↓ Cdx2, Ets2, Sox2 | Loss of TS morphology | Knockdown efficiency poor by qRT-PCR | Kidder and Palmer., 2010 |
| | shRNA knockdown | Stable transfection | ↓ Cdx2, Elf5, Eomes | Visibly differentiating | Efficient knockdown | Gao et al., 2019 |
| | Overexpression | Not specified, but long-term | ↑ Cdx2 and TGC markers ↓ Syna, Synb - inconsistent between figures<br>↑ Eomes, Tfap2c, Hand1<br>↓ Elf5, Gcm1, Syna, Synb | Normal TSC morphology with increased differentiation | RNA-seq performed (no gene list provided) | Latos et al., 2015 |

**Table 1.2: Summary of experiments that have altered stem cell marker gene expression in TSCs.** This table summarises the timeframe, affected genes and morphology after downregulating (yellow) or upregulating (purple) known TSC marker genes. When outlined times contain a '?', the timeframe has been inferred from figures rather than outlined in the text.

## 1.5 *In vitro* differentiation into trophoblast lineages

TSC marker genes are only co-expressed post-implantation, indicating that the *in vivo* equivalents of TSCs are trophoblast stem cells in the ExE of post-implantation embryos (Figure 1.2). However, this has been challenged by recent work suggesting that TSCs are heterogeneous populations, containing cells that represent different developmental stages: from polar trophectoderm to ExE and differentiating cells (Frias-Aldeguer et al., 2019). *In vitro*, TSCs differentiate when key growth factors (FGF4 and TGFβ) that maintain trophoblast stemness are removed (Tanaka et al., 1998). The absence of FGF4 and TGFβ *in vitro* results in spontaneous differentiation to all trophoblast lineages, but is biased towards the TGC lineage and away from Syn populations (Hughes et al., 2004, Latos and Hemberger, 2016). *In vivo,* both NODAL and FGF4 are secreted from the epiblast and are critical for maintaining extraembryonic ectoderm identity (Guzman-Ayala et al., 2004). FGF4 also represses ectoplacental cone identity (Guzman-Ayala et al., 2004). If the ExE is removed from NODAL and FGF4 signalling, this stem cell population differentiates into TGCs (Guzman-Ayala et al., 2004). Therefore, *in vivo* ExE and *in vitro* TSCs primarily differentiate through an EPC-like fate and subsequently into the two EPC derivatives: the spongiotrophoblast and TGCs.

*In vitro*, increased differentiation into Syn derivatives can be driven by removing FGF4 and TGFβ and supplementing with activin (Natale et al., 2009). In addition, generation of Syn derivatives can also be increased by changing components of the extracellular matrix or reducing histone deacetylase (*Hdac*) activity alongside growth factor withdrawal (Choi et al., 2013, Maltepe et al., 2005). Recent work by Murray et al (2016) found that the relative expression of the cell surface protein PLET1 is indicative of cell fate during differentiation. Trophoblast cells that express high levels of PLET1 on the surface preferentially differentiate into TGCs. Low or absent PLET1 is indicative that cells will differentiate into Syn derivatives. As a cell surface protein, the differential expression of PLET1 could be used in future experiments to enrich for EPC-like TGC progenitors or chorion/labyrinth-like Syn progenitors (Murray et al., 2016).

## 1.6 Beyond the placenta: *Cdx2* in mouse embryogenesis

In addition to the trophoblast lineage, *Cdx2* is required for the normal development of multiple other embryonic lineages. The role that *Cdx2* plays in other lineages has been extensively studied. Understanding how *Cdx2* behaves in different lineages may help to understand the role of *Cdx2* in the trophoblast lineage and/or may uncover universal roles of *Cdx2* across tissues.

*Cdx2* is one of three mouse paralogues of the Drosophila gene caudal (cad): *Cdx1* (Duprey et al., 1988), *Cdx2* (James and Kazenwadel, 1991) and *Cdx4* (Gamer and Wright, 1993). All *Cdx* genes are members of the ParaHox homeobox gene family (Ferrier et al., 2005). Unlike *Cdx2*, *Cdx1* and *Cdx4* are not expressed in preimplantation embryos or the trophoblast lineage (Meyer and Gruss, 1993, Gamer and Wright, 1993, Beck et al., 1995, Strumpf et al., 2005).

All *Cdx* genes are expressed in the three germ layers of the late primitive streak at E7.2-7.5, and both *Cdx2* and *Cdx4* are expressed in the allantois (Meyer and Gruss, 1993, Beck et al., 1995, Gamer and Wright, 1993). *Cdx2* expression in the allantois is essential to its function as tetraploid aggregation rescued *Cdx2*-null embryos die at E11.5 due to failed chorio-allantoic fusion (Chawengsaksophak et al., 2004). By E8.5, *Cdx1, 2* and *4* are expressed in the neural tube, mesoderm of the tail bud and the neuromesodermal progenitor population that drives axial elongation (Tsakiridis et al., 2014, Gouti et al., 2017). Additionally, *Cdx2* and *Cdx4* are expressed in the presomitic mesoderm (Beck et al., 1995, Gamer and Wright, 1993).

*Cdx4*-null embryos are viable and show no phenotype (van Nes et al., 2006). *Cdx1*-null embryos, although viable and fertile, show defects in the anterior cervical vertebrae (Subramanian et al., 1995). In contrast, rescued *Cdx2*-null embryos are severely truncated, generating a maximum of 17 posteriorly deformed somites (Chawengsaksophak et al., 2004). Not only is *Cdx2* expression important for axial elongation, but it also needs to be present at wild-type levels as *Cdx2*-heterozygote mice show anterior deformities and shortened tails (Chawengsaksophak et al., 1997). Double knockout *Cdx2/4* embryos are more severely affected than *Cdx2*-null embryos (Young et al., 2009). *Cdx1/2/4* triple knockout embryos are more severely

affected still, failing to generate any post-occipital tissue (van Rooijen et al., 2012). In all, this suggests that *Cdx* genes can compensate for one another to a certain extent, but *Cdx1* and *Cdx4* are poorer at compensating for *Cdx2* than vice versa.

After birth, the only tissues *Cdx2* is expressed in are the small and large intestine, where *Cdx1* is also expressed (Duprey et al., 1988, James and Kazenwadel, 1991, Beck et al., 1995). All three *Cdx* genes are expressed in the developing hindgut (Meyer and Gruss, 1993, Beck et al., 1995, Gamer and Wright, 1993). Whilst neither *Cdx1*-null nor *Cdx4*-null mice show intestinal defects, the caecum of *Cdx2*-heterozygotes has regions that are incorrectly specified (Chawengsaksophak et al., 1997). Similarly, if *Cdx2* is deleted before the intestine is specified, the intestinal epithelium undergoes an oesophageal-like conversion (Gao et al., 2009). In contrast, if *Cdx2* expression is lost after the intestine is specified, this compromises intestinal proliferation, differentiation and overall function (Verzi et al., 2010, San Roman et al., 2015).

These phenotypic differences are likely caused by CDX2 binding and activating different target genes at different stages of intestinal development (Kumar et al., 2019). Those sites that CDX2 binds to at these different intestinal stages are already accessible when CDX2 binds there, indicating that CDX2 responds to chromatin changes rather than establishing them (Kumar et al., 2019). Similarly, when overexpressed in various cell types, CDX2 binds to regions of accessible chromatin rather than altering the chromatin landscape (Mahony et al., 2014). This is in direct contrast to recent work investigating neural progenitor commitment during axis elongation. Metzis et al (2018) instead suggest that CDX transcription factors are required to remodel chromatin at a specific development time window (Metzis et al., 2018).

## 1.7 Aims

To date, the literature indicates that *Cdx2* has independent roles in different lineages. In addition, *Cdx2* has also been shown to have multiple roles in the same lineage at different developmental stages. It is well established that *Cdx2* is important in the early trophoblast context for repressing the pluripotency network during trophectoderm establishment. However, *Cdx2* expression is required in trophoblast stem cell populations long after this role becomes obsolete. What, if any, is its direct function later in this lineage?

The aim of this thesis is to establish both the immediate and long-term effects of *Cdx2* loss on the established trophoblast fate. This work will provide new insights into the differential roles *Cdx2* may play at different trophoblast developmental stages. Furthermore, this analysis will unpick the role of *Cdx2* in the gene regulatory network that maintains the trophoblast fate.

- In Chapter 3, I will determine CDX2 binding sites in established trophoblast stem cells and address the possibility that CDX2 binding differs before and after trophoblast fate is established.

- In Chapter 4, I will examine the effect of transient *Cdx2* loss on the transcriptome and the chromatin landscape of TSCs.

- In Chapter 5, I will define the effect of sustained *Cdx2* loss on the TSC fate. In addition, I will validate unique enhancers that may be responsible for observed phenotypes.

- In Chapter 6, I will further validate the phenotype caused by long-term *Cdx2* loss and its effect on the TSC gene regulatory network by deleting *Cdx2* using CRISPR-Cas9 genome editing.

# Chapter 2. Materials & Methods

For all data in this thesis, except those generated during bioinformatic analyses (see 2.5), data was analysed in Microsoft Excel (Microsoft: Washington, USA) and graphs generated and statistical analyses performed by Prism8 (Graphpad: La Jolla, USA).

## 2.1 Cell Culture

All cell culture was performed on Costar® plastic (Corning, USA).

### 2.1.1 Mouse trophoblast stem cells

Three mouse trophoblast stem cell (TSC) lines were used in this thesis. INX and XYO cell lines were derived by Daniel Snell (Turner Lab, The Francis Crick Institute, UK). Rossant-GFP TSCs were a gift from Janet Rossant (Hospital for Sick Children, Canada).

#### *2.1.1.1 Cell maintenance*

Cells were maintained as previously described (Tanaka et al., 1998, Hayakawa et al., 2015). Briefly, cultures were maintained in TSC maintenance medium (TSCMM) (Appendix A) and incubated at 37°C in 5% $CO_2$. TSCs were passaged as follows: cells were washed briefly in PBS, incubated in TrypLE (12604013, Gibco) at 37°C for five minutes, disaggregated by pipetting and quenched with 1.5 mL TSCM (Appendix A). Cells were passaged 1:20 approximately every four days with medium changes every two days. For cell density experiments, cells were plated at 1.05 x $10_4$ cells per $cm_2$. Cell counting was performed using an EVE Automatic Cell Counter (NanoEnTek Inc., Korea). For ChIP-seq experiments, TSCs were plated in 100 mm dishes and harvested at approximately 70% confluency. Cells were subjected to in-house *Mycoplasma* screening every two months (Cell Services Science Technology Platform, The Francis Crick Institute).

#### *2.1.1.2 Transfection*

Multiple reagents, vector concentration and incubation times were trialled when optimising transfection efficiency. As previously shown, jetPRIME was the most

efficient reagent (Hayakawa et al., 2015) and the maximum transfection efficiency was approximately 30%.

The transfection reagent jetPRIME (114-07, Polyplus-transfection®) was used to deliver the vectors for shRNA knockdown, CRISPR-Cas9 genome editing and enhancer validation experiments. For shRNA and CRISPR-Cas9 experiments, cells were passaged as described above (see 2.1.1.1), plated at 7 x $10^4$ cells per 6 well plate well in 1.5 mL TSCMM (Appendix A) and left to adhere overnight (approximately 16 hours). For each 6 well plate well transfected, 6 µg of plasmid was diluted in 400 µL of jetPRIME buffer. Subsequently, 15 µL of jetPRIME reagent was added, mixed briefly and incubated for 10 minutes at room temperature. The transfection mix was added to the existing medium on the cells and incubated four hours; after which the medium was aspirated and fresh TSCMM was added. For shRNA knockdown experiments, cells were co-transfected with 2 µg each for three vectors (see 2.2.7.2). For CRISPR-Cas9 co-transfections using two vectors, 3 µg of each vector (see 2.2.7.4) was used.

For dual-luciferase assay experiments, 1.75 x $10^4$ cells were plated per 24 well plate well in 300 µL TSCMM (Appendix A) and incubated for 16 hours to adhere. The activity of enhancer constructs (see 2.2.7.3) was assayed in both the presence (+GF) and absence (-GF) of growth factors. For this, six 24 well plate wells were plated per enhancer construct per experiment: three biological replicates for +GF and -GF, respectively. Immediately before transfection, the medium on -GF wells was replaced with fresh TSCM only (Appendix A). A total of 1080 ng of firefly luciferase construct and 120 ng of pRL Renilla Luciferase Control Reporter Vector (E2231, Promega) was diluted in 100 µL jetPRIME buffer and 3 µL jetPRIME reagent added for each 24 well plate well. The transfection mix was vortexed, incubated at room temperature for ten minutes, added to cells and incubated for four hours. Subsequently, the medium was changed to TSCMM and TSCM for +GF and -GF conditions, respectively.

#### *2.1.1.3 Drug selection for transfected cells*

To select for positively transfected cells, 0.7µg/mL puromycin (A1113803, Thermo Fisher Scientific) was added to TSCMM (Appendix A) and used to replace the medium on cells 48 hours post-transfection. Cells were incubated in puromycin for 32 hours, before the medium was aspirated. Cells were washed in PBS and cultured thereafter in TSCMM.

#### *2.1.1.4 Differentiation*

To induce spontaneous differentiation in TSCs, cells were plated at 1 x $10_4$ cells per 6 well plate well in TSCMM (Appendix A) for 16 hours. Concurrent with the start of transfection experiments, the medium was changed to TSCM only (Appendix A) which is devoid of any growth factors. Cells were collected for immunofluorescence (see 2.3.1) and RNA isolation (see 2.2.2) in parallel with transfection experiments.

#### *2.1.1.5 Fluorescence-activated cell sorting (FACS)*

For shRNA knockdown and CRISPR-Cas9 genome editing experiments, constitutively active fluorescent protein cassettes were used to visualise and isolate transfected cells. Cells were isolated in bulk for both knockdown and genome editing experiments. For CRISPR-Cas9 genome editing, single-cell sorting was also performed to generate clonal cell lines. Due to high demand, a range of cell sorting machines available in the Flow Cytometry Science Technology Platform at The Francis Crick Institute were used, including: MoFlo XDP (Beckman Coulter, USA), BD FACSAria™ Fusion (BD Biosciences, USA) and Avalon (Bio-Rad Laboratories, USA) cell sorters. Cells were dissociated as described (see 2.1.1.1), pushed through cell strainer lids (352235, Corning, Fisher Scientific) on polypropylene 5 mL FACS tubes (352063, Corning, Fisher Scientific) and centrifuged for three minutes at 200x *g*. The supernatant was then aspirated and the pellet re-suspended in approximately 100 µL PBS containing 1% foetal bovine serum (FBS, 16000044, Thermo Fisher Scientific) per 6 well plate well. Tube were kept on ice for the entire sort. For ATAC-seq, 50,000 cells were collected in 1.5 mL tubes (0030108051, Eppendorf) containing PBS. For collections to isolate RNA, cells were sorted directly into TRIzol (15596018, Thermo Fisher Scientific).

#### *2.1.1.6 Cryopreservation*

TSCs were frozen in TSCM (Appendix A) containing 10% dimethyl sulfoxide (DMSO, D2650, Merck). Approximately six cyrovials (363401, Thermo Fisher Scientific) containing 1 mL of cells each was obtained from a 6 well plate well. Vials were placed at -80°C overnight in a cell freezing container and were subsequently transferred to permanent storage in liquid nitrogen.

### 2.1.2 Mouse embryonic fibroblasts

Vials of frozen P0 MF1 mouse embryonic fibroblasts (MEFs) (1 x confluent T175) were supplied by the Human and Embryo Stem Cell Unit (HESCU, The Francis Crick Institute).

#### *2.1.2.1 Production of conditioned medium for TSCs*

Cells were thawed into a 150 mm dishes and cultured in MEF medium (Appendix A) at 37°C in 5% $CO_2$. For passaging, cells were dissociated using trypsin-EDTA (25300054, Gibco, Thermo Fisher Scientific) and were expanded for three passages to generate in excess of 10 x $10^7$ cells. After dissociation, cells were exposed to an approximate 35 Gy dose of radiation using a caesium source. Thereafter, cells were counted and plated at 6 x $10^6$ cells per 150 mm plate in 27 mL TSCM (Appendix A). Every three days, for a total of five collections, conditioned medium (CM) was collected, frozen and replaced with fresh TSCM. Once all collections were completed, batches were pooled, centrifuged and filtered to remove any MEF cell debris. Filtered conditioned medium was stored at -20°C.

## 2.2 Molecular Biology

Primers were either obtained from published work or designed using Primer3 (Untergasser et al., 2007). Primer specificity was confirmed using primer-BLAST (Ye et al., 2012). The sequences of all oligonucleotides used within this thesis are provided in Appendix B (ordered from Sigma-Aldrich/Merck, USA).

### 2.2.1 Genomic DNA extraction

Genomic DNA (gDNA) was extracted from TSCs for PCR amplification of enhancer constructs and to genotype CRISPR-Cas9 genome edited TSCs. After dissociation, (see 2.1.1.1) cells were centrifuged at 200x *g* for three minutes. Pellets were resuspended in KT buffer (Appendix A) according to the number of cells isolated (approximately 10 µL per 1 x $10_4$ cells) and incubated at 55°C for 60 minutes and at 95°C for 15 minutes. Lysates were stored at -20°C.

### 2.2.2 RNA extraction

For CRISPR-Cas9 genome edited clones, RNA was extracted using the RNeasy® Mini Kit (74106, QIAGEN) according to manufacturer's instructions (Genomics Equipment Park, The Francis Crick Institute). RNA was isolated for RNA-seq and shRNA knockdown experiments as follows: cells were washed in PBS, 500 µL TRIzol was added (15596018, Thermo Fisher Scientific) and transferred to 1.5 mL LoBind Eppendorf tubes (0030108051, Eppendorf). Tubes were snap-frozen and stored at -80°C. After thawing on ice, 100 µL chloroform (C2432, Sigma-Aldrich, Merck) was added per 500 µL TRIzol, mixed and transferred to pre-spun phase-lock tubes (733-2478, Quanta Bio, VWR). Samples were centrifuged at 12,000 rpm for 30 minutes at 4°C. The upper aqueous phase containing RNA was transferred to a new RNase-free tube (0030108051, Eppendorf) and an equal volume of 70% ethanol added ((32221, Sigma-Aldrich, Merck), diluted in DNase-free water). Samples were incubated for 10 minutes at room temperature, transferred to RNeasy MinElute Spin Columns (74034, QIAGEN) and centrifuged for 30 seconds at 10,000 rpm. Buffer RW1, Buffer RPE and 80% ethanol ((32221, Sigma-Aldrich, Merck), diluted in DNase-free water) were applied in subsequent wash steps, centrifuged and discarded. After re-centrifugation to remove any residual wash buffer, RNA was eluted in 14 µL RNase-free water. RNA concentration was quantified using a Nanodrop spectrophotometer (Thermo Fisher: Waltham, USA).

### 2.2.3 cDNA synthesis

RNA was isolated as described (2.2.2) and converted to cDNA using the Maxima First Strand cDNA Synthesis Kit according to the manufacturer's protocol (K1672, Thermo Scientific). For each cDNA synthesis reaction, between 250 ng and 1 µg

total RNA was used. cDNA dilutions were calculated as 0.6x RNA concentration (ng) of $dH_2O$ (µL) minus the total final volume of the cDNA synthesis reaction (20µL). Minus reverse transcription controls were generated to detect the presence of any contaminating genomic DNA.

### 2.2.4 Polymerase chain reaction (PCR)

For all cloning, sequencing and MiSeq genotyping reactions, a high-fidelity polymerase was used (M0493S, NEB) as per the manufacturer's protocol. For PCR primer validation, colony PCR and TA cloning, KAPA2G Fast ReadyMix was used (KK5103, Kapa Biosystems) according to the manufacturer's protocol. All PCR reactions used an appropriate volume of DNA relative to its concentration. Primers were used at a final concentration of 0.2 µM (Sigma-Aldrich/Merck, USA). Standard PCR reaction volumes were between 20 µL and 50 µL depending on the intended subsequent use. PCR reactions were performed in a thermocycler (T100 Thermal Cycler, BioRad). To visualise DNA fragment lengths, PCR products were loaded onto an 1-2% agarose/TAE gel (16500500, Invitrogen) containing 1:50,000 RedSafe (21141, Ecogen) with an appropriately sized DNA ladder (NEB). Gel electrophoresis was performed at around 100 V for an appropriate length of time to resolve bands and images were acquired on the ChemiDoc™ XRS+ Imaging System (Bio-Rad Laboratories, USA) using proprietary Quantity One software (Bio-Rad Laboratories, USA).

### 2.2.5 Quantitative real-time PCR (qRT-PCR)

All primers were tested for their specificity using PCR on cDNA followed by gel electrophoresis to assess the fragment length and number of products. qRT-PCR analysis was performed using the established 2^(-delta delta CT) method (Livak and Schmittgen, 2001). As such, primers were tested using standard curve analysis. If their efficiency was $2 \pm 0.1$, this was deemed to be sufficient. RNA was extracted (see 2.2.2) and converted to cDNA (see 2.2.3) as described. As per the manufacturer's instruction, 384 plates (4ti-0380/C, 4titude) were prepared with Lightcycler 480 SYBR Green I master mix (04887352001, Roche). Each biological replicated was plated in technical triplicate and run on the Lightcycler 480 II thermal

cycler (Roche, Switzerland). Data were exported from proprietary Roche software and processed in Excel (Microsoft, USA).

### 2.2.6 Solid phase reverse immobilisation (SPRI) bead purification

SPRI beads were generated and validated in-house by Rita Monteiro (Smith Lab, The Francis Crick Institute). For ATAC-seq and MiSeq, beads were applied at ratios of 1:1.5 and 1:1 per volume of DNA, respectively, in a 96 well microplate (4ti-0117, 4titude). Bead-DNA samples were left to incubate for five minutes and then placed on a magnetic stand (AM10027, Invitrogen, Thermo Fisher Scientific) for five minutes, allowing beads to separate. Bead-DNA complexes were washed twice in 80% ethanol ((32221, Sigma-Aldrich, Merck), diluted in DNase-free water) and the supernatant discarded. Microplates were sealed using an adhesive PCR film (4ti-0500, 4titude) and centrifuged at 200x *g* for one minute. Microplates were replaced on the magnetic stand and all remaining supernatant removed. After airdrying on the magnetic stand, microplates were removed from this stand and beads were re-suspended in 20 µL 10 mM Tris-HCl to elute DNA. After a five-minute incubation, the microplate was re-placed on the magnetic stand, the eluate collected and stored at -20°C.

### 2.2.7 Cloning of vectors

Once cloning of each vector was confirmed, large quantities of plasmids were generated using ZymoPURE II Plasmid Maxiprep Kits (D4202, Zymo Research).

#### *2.2.7.1 Colony PCR*

Individual colonies were isolated in 50 µL Buffer TE (Media Preparation, The Francis Crick Institute). PCR reactions were performed using 2 µL of colony in TE, 0.2 µM final concentration of primers and 12 µL 2X KAPA2G Fast ReadyMix (KK5103, KAPA Biosystems, Roche). PCR conditions were as follows: pre-incubation at 98°C for 3 minutes, 35 cycles of; 98°C for 15 seconds, 60°C for 15 seconds, 72°C for 1 minute; final extension at 72°C for 2 minutes. Extension time (72°C) within the cycles varied depending on the size of the intended fragment from 15 seconds to two minutes. 20 µL PCR product was used to visualise and verify fragment size in 1-2% agarose/TAE gel containing 1:50,000 RedSafe (21141, Ecogen). Positive bacterial clones

containing the desired size insert were subsequently used to generate mini preps, extracted using the QIAprep Spin Miniprep Kit (27106, QIAGEN), and the insert was sequenced by sanger sequencing (Genomics Equipment Park, The Francis Crick Institute).

#### *2.2.7.2 pLKO.1-mCherry-IRES-puromycin*

Five TRC (The RNAi Consortium) Lentiviral shRNA vectors that target *Cdx2* were purchased from Dharmacon-GE (Horizon, UK) (Table 2.1).

| TRC Clone ID | Thesis name | Target Sequence | Vector |
|---|---|---|---|
| TRCN0000055393 | shRNA1 | GTTTCACTTTAGTCGATACAT | pLKO.1 |
| TRCN0000055394 | shRNA2 | GCTCTCCGAGAGGCAGGTTAA | pLKO.1 |
| TRCN0000055395 | shRNA3 | CCGCATCATCACCCGCACCAT | pLKO.1 |
| TRCN0000055396 | shRNA4 | CGGGTGGTGTACACAGACCAT | pLKO.1 |
| TRCN0000055397 | shRNA5 | GCGAAACCTGTGCGAGTGGAT | pLKO.1 |
| N/A | SCR | CCTAAGGTTAAGTCGCCCTCG | pLKO.1 |

**Table 2.1: Purchased shRNA vectors targeting Cdx2.**

Existing puromycin cassettes in pLKO.1 vectors were replaced by nuclear mCherry-IRES-puromycin (mCIP) amplified from pBRY-nuclear mCherry-IRES-PURO (primers in Appendix B, Merck). pBRY-nuclear mCherry-IRES-PURO was a gift from Jacob Hanna (Addgene plasmid # 52409 ; http://n2t.net/addgene:52409 ; RRID:Addgene_52409). PCR conditions were as follows: pre-incubation at 98°C for 3 minutes, 35 cycles of; 98°C for 20 seconds, 68°C for 30 seconds, 72°C for 150 seconds; final extension at 72°C for 2 minutes (see 2.2.4). pLKO.1 vectors were linearised using KnpI-HF (R3142S, NEB) and BamHI-HF (R3136S, NEB) in CutSmart® Buffer at 37°C for one hour. Linearised pLKO.1 vectors and mCIP PCR products were resolved by gel electrophoresis, the correct band cut out and purified using the QIAquick Gel Extraction Kit (28706, QIAGEN) as per the manufacturer's protocol. mCIP PCR products were cloned in to each pLKO.1 vector using In-Fusion® HD Cloning Kit (638920, Takara Bio) according to the manufacturer's protocol. Ligated plasmids were transformed into TOP10 competent bacteria (C4040-06, Thermo Fisher Scientific), plated on agar infused with ampicillin (Media Preparation, The Francis Crick Institute) and incubated at 37°C overnight. Colony PCR was performed on 6-8 colonies as described (see 2.2.7.1) using mCIP_F and mCIP_R primers (Appendix B).

#### *2.2.7.3 Enhancer validation constructs (pGL3-basic)*

Several genomic regions shown to increase in accessibility under *Cdx2* knockdown conditions around the *Hand1* locus were assessed for their enhancer activity on the *Hand1* promoter. To perform enhancer validation experiments, pGL3 Luciferase Reporter vectors (E1751/E1771/E1761, Promega) containing a modified firefly luciferase gene were used in a dual luciferase assay.

Three *Hand1* promoters of differing lengths (209bp, 283bp and 673bp, see Figure 5.11a) were tested for their efficiency (primers in Appendix B). These were PCR amplified from INX gDNA (see 2.2.1) using Q5® High-Fidelity DNA Polymerase (M0493S, NEB) according to the manufacturers protocol. PCR conditions were as follows: pre-incubation at 98°C for three minutes, 40 cycles of; 98°C for 15 seconds, 60°C for 15 seconds, 72°C for 30 seconds; final extension at 72°C for 1 minutes. Fragments were purified using the QIAquick PCR purification kit (28104, QIAGEN) as per the manufacturer's protocol. Restriction digest reactions were performed for each promoter fragment, the pGL3-basic vector (E1751, Promega) and the pGL3-enhancer vector (E1771, Promega) using XhoI (R0146S, NEB) and HindIII-HF (R3104S, NEB) in CutSmart® buffer at 37°C for one hour. Linearised pGL3-basic vector (500ng) were dephosphorylated and the ligation of PCR fragments and vectors performed using the Rapid DNA Dephosphorylation & Ligation Kit (04-898-125-001, Roche) as per the manufacturer's protocol. Ligated plasmids were transformed into TOP10 competent bacteria (C4040-06, Thermo Fisher Scientific) and selected on ampicillin-infused agar plates. Colony PCR was performed on 6-8 colonies (see 2.2.7.1) using pGL2-Basic_F and LucNrev primers (Appendix B). Three clones with the desired insert were sanger sequenced for each new vector. Any deviations in sequence from the reference GRCm38.p6 Ensembl genome had to be consistent between pGL3-basic and pGL3-enhancer and are highlighted in Appendix C

Although all three promoters tested are active and respond to the SV40 enhancer, the 283bp promoter had the most similar baseline activity to the SV40 promoter only control vector (E1761, Promega). As such, this promoter was chosen to assay the enhancer capabilities of six (Region A-F) differentially accessible regions found

around the *Hand1* locus under shRNA knockdown conditions. Two control regions, C1 and C2, representing inaccessible and non-changing accessible regions, respectively, were also cloned into the 283bp *Hand1* promoter pGL3-basic vector. Primers were ordered with SalI restriction sites added to the 5' end of both oligonucleotides (Appendix B, Merck). Each enhancer region was amplified and PCR purified using the same protocol as described for the promoter regions above. Restriction digest reactions for SalI-HF (R3138S, NEB) were performed on all amplicons and the 283bp Hand1 promoter pGL3-basic vector according the manufacturer's protocol. The linearised plasmid was dephosphorylated and ligated with the enhancer constructs using the Rapid DNA Dephosphorylation & Ligation Kit (04-898-125-001, Roche) following the manufacturer's protocol and transformed into TOP10 bacteria (C4040-06, Thermo Fisher Scientific). As only one restriction enzyme was used, the enhancer regions could insert in two directions. Colony PCR was performed for eight colonies on ampicillin-infused agar plates as described (see 2.2.7.1) using RVprimer4 and specific forward primers to the respective enhancer (Appendix B), enabling the selection of colonies that contained the insert in the desired orientation.

Enhancer B and D proved difficult to clone by conventional cloning. As such, both were cloned into dephosphorylated SalI-linearised plasmids (primers in Appendix B) using the In-Fusion® HD Cloning Kit (638920, Takara Bio) according to the manufacturer's protocol. In-Fusion® products were transformed into TOP10 bacteria (C4040-06, Thermo Fisher Scientific) and several colonies on ampicillin-infused agar plates were subject to colony PCR (see 2.2.7.1).

Sanger sequencing (Genomics Equipment Park, The Francis Crick Institute) was performed on three colonies with the desired orientation for all enhancers to validate the sequence and characterise any deviations from the reference GRCm38.p6 Ensembl genome. Plasmids showing minimal number of deviations from this reference genome were selected for dual luciferase assays. Any characterised deviations in sequence from the reference genome are outlined in Appendix C.

#### *2.2.7.4 px330-mCherry-gRNA*

Guide RNA (gRNA) sequences were designed using the MIT CRISPR Designer tool (Cong et al., 2013, Ran et al., 2013). BbsI restriction sites were added to the 5' end of each oligonucleotide for the six chosen gRNAs (G2a-f, Merck, Appendix B). Oligonucleotide pairs were annealed and cloned into the px330-mCherry plasmid, a gift from Jinsong Li (Addgene plasmid # 98750; http://n2t.net/addgene:98750; RRID:Addgene_98750), using an established digestion-ligation protocol (Cong et al., 2013). Each ligation reaction was transformed into TOP10 competent bacteria (C4040-06, Thermo Fisher Scientific), plated on agar infused with ampicillin (Media preparation, The Francis Crick Institute) and incubated at 37°C overnight. Colony PCR (see 2.2.7.1) was performed on six clones for each gRNA inserted using hU6 and the reverse oligonucleotide for each G2a-f gRNA (primers in Appendix B) and visualised on a 1.5% agarose/TAE gel containing 1:50,000 RedSafe (21141, Ecogen). Three clones per gRNA were used to generate mini preps and eluted using the QIAprep Spin Miniprep Kit (27106, QIAGEN) as per the manufacturer's protocol. Correct orientation and sequence were confirmed by sanger sequencing using the hU6 primer (Appendix B) (Genomics Equipment Park, The Francis Crick Institute).

#### *2.2.7.5 TA cloning*

To validate the genotype of *Cdx2* clones, full-length *Cdx2* transcripts were PCR amplified using 5to3_Cdx2 primers (Appendix B) using 2X KAPA2G Fast ReadyMix (KK5103, KAPA Biosystems, Roche) according to the manufacturer's protocol (see 2.2.4). PCR products, purified using the QIAquick PCR purification kit (28104, QIAGEN) as per the manufacturer's instructions, were TA cloned into the pGEM®-T Easy Vector System (A1360, Promega). TOP10 competent bacteria (C4040-06, Thermo Fisher Scientific) were transformed and plated on ampicillin infused agar plates. Mini preps were generated for up to 15 colonies and the vectors purified and sanger sequenced using the 5to3_Cdx2 primers by the Genomics Equipment Park at The Francis Crick Institute.

### 2.2.8 Flow cytometry

Flow cytometry was performed to quantify the expression of CDX2, GATA3, HAND1 and TFAP2C in wild-type, shRNA knockdown and CRISPR-Cas9 genome editing

experiments. Cells were dissociated as described (see 2.1.1.1), filtered using cell strainers on FACS tube (352235, Corning, Fisher Scientific) and pelleted for three minutes at 200x *g*. Pellets were re-suspended in cold 4% PFA (28908, Thermo Fisher Scientific) whilst vortexing cells to reduce cell clumping. Cells were fixed for 15 minutes on a rotator. To wash cells, tubes were filled with PBS, centrifuged at 250x *g* for three minutes and the supernatant discarded three times. To permeabilise and block, cells were pelleted for three minutes at 300x *g* and re-suspended in blocking buffer (1% BSA (A9647, Sigma-Aldrich, Merck), 0.5% Tween-20 (NAT1082, Scientific Laboratory Supplies)) and incubated for one hour at room temperature. Cells were counted using the EVE automatic cell counter (NanoEnTek Inc, Korea).

For staining, antibodies were diluted in blocking buffer as follows: 1:5 anti-CDX2 (560395, BD Biosciences), 1:5 anti-GATA3 (560405, BD Biosciences), 1:75 anti-HAND1 (AF3168, R&D Systems) and 1:50 anti-TFAP2C (AF5059, R&D Systems). In all cases, 10 µL antibody suspension was added per 1 x $10^5$ cells up to a maximum total of 1 x $10^6$ cells (equivalent to 100 µL antibody). Antibody-cell suspensions were incubated for 30 minutes at room temperature in the dark. Unstained and single antibody stained cells were generated for each antibody in every experiment. Cells were washed in three times in PBS, centrifuging at 250x *g* for three minutes. For the unconjugated anti-HAND1 and anti-TFAP2C antibodies, cells were stained with donkey anti-goat Alexa fluor 647 (A-21447, Invitrogen, Thermo Fisher Scientific) diluted 1:500 in blocking buffer. As with the primary antibodies, 10 µL antibody suspension was added per 1 x $10^5$ cells and incubated for 30 minutes at room temperature. Cells were washed three times in PBS by filling the tube, centrifuging for three minutes at 200x *g* and discarding the supernatant. Cell pellets were resuspended in 200 µL PBS per condition and 10 µL of 250 µg/µL DAPI (DAPI (40011, Biotium), diluted to 5 mg/mL in PBS) were added to each experimental condition and DAPI only controls.

Flow cytometry analysis was performed on one of the BD LSR Fortessa or BD LSR II (BD Biosciences, USA) machines available in the Flow Cytometry Science Technology Platform (The Francis Crick Institute). Analysis of flow cytometry data was performed on FlowJo™ 10.3 (BD Biosciences, USA).

### 2.2.9 Dual luciferase assay for enhancer validation

All enhancer validation assays were performed in TSCs. Transfections were performed as described (see 2.1.1.2). Dual luciferase assay was performed using the Dual-Luciferase® Reporter Assay System (E1910, Promega). Luciferase Reagent (LARII) and Stop & Glo® reagents were diluted as per the manufacturer's recommendations and then 1:10 with $dH_2O$. LARII and Stop & Glo® reagents were stored at -20°C. At approximately 24-hours post-transfection, wells were washed in PBS and lysed in 150 µL 1 x passive lysis buffer (PLB) for 30 minutes at room temperature in the dark on a shaking plate. 5 µL lysate was transferred per sample in technical duplicate and measured one at a time on the microplate luminometer (Centro XS3 LB960, Berthold Technologies) using the following conditions: 100 µL of LARII was added to a well, there was a delay of two seconds and the luminometer measured firefly (pGL3) activity. Subsequently, the luminometer added 100 µL of Stop & Glo® reagent, delayed for two seconds and measured *Renilla* luciferase (pRL) activity. MikroWin software (Labsis, Germany) was used to generate ratios between firefly and *Renilla* luciferase and data was exported and processed in Excel (Microsoft, USA).

### 2.2.10 Clustered regularly interspaced short palindromic repeats (CRISPR) mutation analysis

Six gRNAs (G2a-f) were cloned into the px330-mCherry vector (Wu et al., 2013) (see 2.2.7.4) and transfected into TSCs alone or in pairs (see 2.1.1.2). Positively transfected cells were isolated by their mCherry expression in bulk or as single-cells (see 2.1.1.5). Genotyping of clonal and polyclonal populations was performed by one of two methods.

#### *2.2.10.1 MiSeq*

Three primer sets were designed to generate PCR fragments of 360-380bp around the targeted region (Chapter 6, Figure 6-6a). Overhang adapter sequences required for further library preparation were added to the 5' end of forward and reverse oligonucleotides (ordered from Merck, Appendix B). gDNA was extracted as described (see 2.2.1) and PCR amplified using Q5® High-Fidelity DNA Polymerase (M0493S, NEB). PCR conditions were as follows: pre-incubation at 98°C for 3

minutes, 30 cycles of; 98°C for 20 seconds, 61.4°C for 20 seconds, 72°C for 30 seconds; final extension at 72°C for 1 minutes. Presence of PCR amplicons was confirmed by running a small aliquot on a 2% agarose/TAE gel with 1:50,000 RedSafe (21141, Ecogen). Amplicons were subsequently purified using a SPRI bead purification (see 2.2.6) and eluted in 20µL 10 mM Tris-HCl. For single-cell clonal analysis, amplicons were generated with one of the three MiSeq primer sets with an amplicon for each primer set pooled together to internally multiplex each sample. Further library preparation using Illumina TruSeq indexing primers (Illumina) and DNA quantification were performed by Matthew Winder (Genetic Modification Service, The Francis Crick Institute). Samples were submitted for sequencing to the Advanced Sequencing Facility (ASF) at The Francis Crick Institute. Following various quality control assessments, libraries were sequenced on the Illumina MiSeq platform and generated 250bp paired-end reads.

For bulk sample analyses, mutations and their frequencies were determined using Cas-Analyzer using the WT marker analysis parameter (Park et al., 2017). Tables were downloaded and subdivided into one of four categories: wild-type and/or substitution (WT/SUB), frameshift, small in-frame or large in-frame mutations using Excel (Microsoft, USA).

For single-cell clonal analysis, paired-end reads were analysed as single-end reads to compare genotypes between reads. Clonal genotyping analysis was performed by Harshil Patel (Bioinformatics, The Francis Crick Institute). Briefly, fastq files were mapped to the mm10 genome independently using bbmap (version 37.00) (Bushnell, 2016) with the parameters “in=<FASTQ_FILE> outm<SAM_FILE> ref=<INDEX> vslow=t maxindel=2000 ambig=toss sam=1.3 mdtag=t nhtag=t xmtag=t amtag=t nmtag=t idtag=t scoretag=t”. Custom scripts using pysam (version 0.9.0) (https://github.com/pysam-developers/pysam) were written to split the mapped reads by the original primer sequence and to calculate indel frequencies. Individual clone genotypes were then inferred using the proportions of reads for each variant within that sample and validated as the same for both paired-end reads.

#### *2.2.10.2 CRISPR cassette screening*

Initial screening of single-cell clones was performed by sanger sequencing, but the prevalence of multiple mutations reduced its usefulness. Briefly, gDNA was PCR amplified using the MiSeq primers (P1, Appendix B) and conditions outlined in 2.2.10.1. The presence of amplicons were confirmed by gel electrophoresis. PCR reactions were enzymatically purified using ExoSAP-IT™ PCR Product Cleanup Reagent (78200.200.UL, Thermo Fisher Scientific) as per the manufacturers protocol by Maria Greco (Genomics Equipment Park, The Francis Crick Institute). Sanger sequencing was performed for each PCR amplicon using the MiSeq primer without the adapter sequence (Genomics Equipment Park, The Francis Crick Institute).

## 2.3 Microscopy

Images were analysed using FIJI (ImageJ, version 2.0.0).

### 2.3.1 Immunofluorescence (IF) staining

TSCs were cultured as described (see 2.1.1.1) on sterilised circular coverslips (Media Preparation, The Francis Crick Institute). Once cells had reached the desired confluency or time point, cell medium was aspirated and cells fixed in ice cold 4% formaldehyde (28908, Thermo Fisher Scientific, diluted in PBS) for 10 minutes at room temperature on a shaking plate. Fixative was removed and wells were washed three times in PBS. Coverslips were stored for up to three months in PBS at 4°C. To permeabilise cells, coverslips were incubated in PBST (0.1% Triton X100 (X100, Merck) diluted in PBS) for 10 minutes at room temperature. To minimise non-specific binding, cells were blocked in blocking buffer (5% donkey serum (34762907, Abcam) diluted in PBST) for one hour at room temperature. Primary antibodies and dilutions used were as follows: rabbit anti-CDX2 1:200 (A300-691A, Bethyl Laboratories), mouse anti-CDX2 1:100 (ab157524, Abcam), anti-TBR2/Eomes 1:250 (ab23345, Abcam), goat anti-HAND1 1:150 (AF3168, R&D Systems), goat anti-TFAP2C 1:150 (AF5059, R&D Systems), goat anti-GATA3 1:200 (AF2605, R&D Systems). All primary antibodies were diluted in blocking buffer and incubated at 4°C overnight in a humidified chamber. Coverslips were washed in PBST for five minutes three times

and subsequently stained with an appropriate AlexaFluor conjugated secondary antibody (Invitrogen, Thermo Fisher Scientific) diluted in blocking buffer at room temperature for one hour. Coverslips were subsequently washed three times with PBST. For long-term *Cdx2* knockdown and *Cdx2* clonal analysis experiments, coverslips were stained for 20 minutes at room temperature with 1:100 Alexa Fluor Phalloidin 488 (A12379, Thermo Fisher Scientific) and 1:10,000 4',6-Diamidino-2-Phenylindole, dihydrochloride (DAPI (40011, Biotium), diluted to 5mg/mL in PBS) diluted in blocking buffer. For all experiments, coverslips were washed in PBST for five minutes three times and mounted onto microscope slides (J1840AMNT, Thermo Fisher Scientific) in VECTASHIELD® Mounting Medium with DAPI (H-1200, Vector Laboratories). High quality images were acquired on the LSM 710 Zeiss Confocal Laser Scanning Microscope using proprietary Zeiss software (Zeiss, Germany).

### 2.3.2 Live cell imaging

Trophoblast stem cells were transfected with SCR or KD (see 2.1.1.2) and selected for using puromycin (see 2.1.1.3). Immediately after, TSCs were placed in an incubator (Okolab, Italy) at 37°C and 5% $CO_2$ on the Nikon Eclipse Ti2 inverted microscope (Nikon, USA). An image was taken at each position every five minutes for 48 hours using Micro-Manager software (version 1.4) (Edelstein et al., 2014).

## 2.4 Next generation sequencing

All quality control metrics and sequencing were performed by the Advanced Sequencing Facility (ASF) at The Francis Crick Institute. Library preparation for all biological replicate samples was performed in parallel to mitigate batch effects.

### 2.4.1 Chromatin immunoprecipitation with massively parallel DNA sequencing (ChIP-seq)

The protocol described below is adapted from a published ChIP-seq protocol optimised in the Smith Lab for early Xenopus embryos (Gentsch and Smith, 2017).

#### *2.4.1.1 Chromatin immunoprecipitation (ChIP)*

INX and XYO TSCs were cultured as described above (section 2.1.1.1) in 3 x 100 mm dishes per ChIP. At approximately 70% confluency, cell medium was aspirated

and adherent cells fixed in fixing buffer (1% formaldehyde (F8775, Merck, diluted in PBS)) for nine minutes on a shaking plate at room temperature. The fixative was aspirated off and cells were washed in ice-cold PBS (Media Preparation, The Francis Crick Institute) containing protease inhibitors (88266, Thermo Fisher Scientific) three times. To detach adherent fixed cells, a cell scrapper was used to isolate cells into 4 mL ice-cold PBS containing protease inhibitors per 100 mm dish. Cells were pelleted at 250x *g* for five minutes at 4°C, snap frozen in liquid nitrogen and stored at -80°C until processing.

To isolate nuclei, pellets were resuspended in 10 mL ice-cold cell lysis buffer (Appendix A) containing protease inhibitors (88666, Thermo Fisher Scientific) and centrifuged at 300x *g* for three minutes. The supernatant was aspirated and pellets resuspended in 5 mL ice-cold CEWB1 (Appendix A) containing protease inhibitors. Cells were homogenised by pipetting and pelleted at 600x *g* for four minutes at 4°C. Pellets were resuspended in 3 mL CEWB1 containing protease inhibitors and incubated for 25 minutes on ice. During this incubation, cells were homogenised by pipetting every 5-7 minutes. Approximately 1 mL of homogenate was transferred to milliTUBE 1 ml AFA Fibre vessels (520135, Covaris) at a time. A Covaris S220 Focused-ultrasonicator (500217, Covaris) was used to sonicate samples using the following settings: duty cycle = 5%, intensity = 4, cycles/burst = 200, procedure time = 420 seconds, temperature = <8°C. Sonicated samples were transferred to 1.5 mL DNA LoBind Tubes (0030108051, Eppendorf) on ice. To remove cellular and nuclear debris, sonicated DNA was centrifuged at 12,000 rpm for 10 minutes at 4°C. The supernatant containing crosslinked DNA transferred to new 1.5 mL DNA LoBind tubes (0030108051, Eppendorf) and stored for up to two days at 4°C. Sonication efficiency was determined before proceeding to chromatin precipitation. For this, 50 µL crosslinked sheared chromatin was added to 50 µL SDS Elution buffer (Appendix A), 5 µL 5M NaCl and 1 µL 20mg/mL Proteinase K (AM2546, Thermo Fisher Scientific) and incubated 65°C for 8-16 hours. De-crosslinked DNA was purified using the QIAquick PCR purification kit (28104, QIAGEN) as per the manufacturer's instructions, eluting twice in 11 µL Buffer EB. After treatment with 0.4 µL RNase A (Invitrogen, 12091-039), fragment sizes were visualised by gel electrophoresis using a 1.4% agarose/TAE gel containing 1:50,000 RedSafe (21141, Ecogen) alongside 1kb (N3232L, NEB) and 100bp (N3231L, NEB) ladders. Shearing was deemed to be

successful if the DNA fragments are distributed between 100-1000bp, with the majority of fragments sized between 300-500bp. Sonication was repeated for 100-200 seconds if shearing was unsuccessful at generating such sized fragments.

Prior to chromatin immunoprecipitation (ChIP) on successfully sonicated samples, approximately 1% of total sheared chromatin volume was transferred to a new DNA lo-bind tube (0030108051, Eppendorf) and stored at 4°C until all ChIP samples were to be de-crosslinked. This sample represents the ChIP input. For this thesis, immunoprecipitation was performed for using a rabbit anti-CDX2 ChIP-grade antibody (A300-691A, Bethyl Laboratories). In total, three x 100 mm dishes were used to generate every library with each 100 mm dish immunoprecipitated using 1.5 µg antibody and incubated overnight on a rotator at 4°C. To wash the ChIP, 60 µL Dynabead Protein G (10003D, Invitrogen) were washed in 1 mL CEWB1 for five minutes on a rotator at 4°C and resuspended in 60 µL CEWB1. In total, 45 µL of washed beads were added to the ChIP sample to account for 4.5 µg total antibody used and incubated for four hours on a rotator at 4°C. To remove unbound and non-specific bound anti-CDX2 antibody, beads were washed ten times with 1 mL pre-chilled RIPA buffer (Appendix A) and once with pre-chilled TEN buffer (Appendix A) for five minutes each. All beads corresponding to a single ChIP experiment were pooled together, centrifuged at 1000x *g* for one minute and placed in a magnetic rack where as much supernatant was removed as possible. Beads were re-suspended in 100 µL SDS elution buffer (Appendix A), incubated at 65°C in a continually vortexing thermomixer for 15 minutes, centrifuged at 15,000g for 30 seconds and the eluate-containing supernatant transferred to a new LoBind Tubes (0030108051, Eppendorf). This step was repeated and the eluates combined.

Input sample volumes were adjusted by adding 180 µL SDS elution buffer to ensure equivalent volumes in both ChIP and input samples for chromatin reverse cross-linking. Both samples were supplemented with 10 µL 5M NaCl, and incubated overnight at 65°C in a hybridisation buffer. After adding 210 µL of TE buffer and 4.2 µL RNase A (Invitrogen, 12091-039), samples were incubated at 37°C for one hour. Subsequently, 4.2 µL 20mg/mL Proteinase K (AM2546, Thermo Fisher Scientific) was added to each sample and incubated at 55°C for two hours. ChIP and input samples were purified using QIAquick PCR purification kit (28104, QIAGEN) as per

the manufacturer's protocol and eluted in 30 µL $dH_2O$. DNA concentration of ChIP and input samples was determined using the Qubit™ dsDNA HS Assay (Q32851, Invitrogen, Thermo Fisher Scientific).

#### *2.4.1.2 Paired-end library preparation*

In total, 10 ng DNA was used to generate each ChIP and Input library using the KAPA Hyper Prep kit (KK8502, Roche). This was performed as per manufacturer's protocol except for the following; Firstly, approximately 1.67 µM of TruSeq DNA adapters (15µM) were used per 10 ng library for Adapter Ligation. Secondly, to convert Y-shaped Illumina adapters to dsDNA, allowing for efficient migration through the agarose gel and reducing the number of adapter dimers, a pre-amplification PCR cycle step was added: 5 cycles of; 98°C for 15 seconds, 60°C for 30 seconds, 72°C for 30 seconds. Finally, after pre-amplification, reactions were cleaned up using 1x volume of SPRI beads and following the protocol outlined in section 2.2.6.

Next, libraries were size selected using an electrophoresis-based method. For this, libraries were run on an E-gel® EX 2% agarose gel (G402002, Thermo Fisher Scientific) using an E-gel® iBase™ Power System (11531226, Fisher Scientific) and DNA between 250-450bp in size cut from the gel. DNA was subsequently purified using the MinElute Gel Extraction Kit (28606, QIAGEN) as per the manufacturer's protocol. Using the HyperPrep kit, DNA was amplified for an appropriate number of cycles in a thermocycler and cleaned-up post-amplification according to the manufacturer's instructions. Library concentration and DNA fragment integrity were determined by Qubit™ dsDNA HS Assay (Q32851, Invitrogen, Thermo Fisher Scientific) and Agilent 2100 Bioanalyzer (Agilent, USA), respectively (Advanced Sequencing Facility, The Francis Crick Institute). ChIP-seq samples were sequenced on the Illumina HiSeq 2500 platform (Advanced Sequencing Facility, The Francis Crick Institute) and generated 101bp paired-end reads.

### 2.4.2 Assay for Transposase Accessible Chromatin with high-throughput sequencing (ATAC-seq)

ATAC-seq libraries were generated using a well-established protocol (Buenrostro et al., 2013, Buenrostro et al., 2015). Briefly, TSCs were cultured and transfected as

described (sections 2.1.1.1 and 2.1.1.2, respectively). For each library generated, 50,000 cells were isolated by FACS (section 2.1.1.5) and maintained on ice. Cells were pelleted at 500x *g* for five minutes at 4°C, re-suspended in 50 µL ice-cold ATAC cell lysis buffer (Appendix A) and homogenised by pipetting. Nuclei extracts, obtained by pelleting cells at 500x *g* for 10 minutes at 4°C, were resuspend in 50 µL transposition reaction mix (Nextera DNA library prep kit, FC-121-1030, Illumina) and incubated at 37°C for one hour. Transposed DNA was purified using QIAquick PCR purification kit (28104, QIAGEN) according to the manufacturer's protocol and eluted 10 µL Buffer EB.

PCR reactions were cleaned-up by SPRI bead purification (section 2.2.6). Prior to sequencing, library concentration and DNA fragment sizes were determined by Qubit™ dsDNA HS Assay (Q32851, Invitrogen, Thermo Fisher Scientific) and Agilent 2100 Bioanalyzer (Agilent, USA), respectively. Sequencing was performed on the Illumina HiSeq 2500 platform (Advanced Sequencing Facility, The Francis Crick Institute) and typically generated ~55 million 51bp paired-end reads per library.

### 2.4.3 RNA-seq

Total RNA was extracted as described (section 2.2.2). A minimum of 500 ng was submitted to the Advanced Sequencing Facility at The Francis Crick Institute whereby samples underwent RNA integrity (RIN) quality control assessment using Caliper LabChip GX (Perkin Elmer, USA). Libraries were generated by the Advanced Sequencing Facility at The Francis Crick Institute from samples that were determined to have a RIN of nine and above using the TruSeq Stranded mRNA Library Prep Kit as per manufacturer's protocol (Illumina, USA). RNA sequencing was carried out on the Illumina HiSeq 4000 platform (Advanced Sequencing Facility, The Francis Crick Institute) and typically generated ~35 million 76bp strand-specific single-end reads per sample.

## 2.5 Bioinformatic analysis

Alignment, generation of visual files, peak calling and differential analysis were performed by Harshil Patel (Bioinformatics, The Francis Crick Institute). Any other work performed by Harshil Patel is indicated.

### 2.5.1 Acquisition of published datasets

Several published datasets were integrated into analyses. All published ATAC-seq data used were downloaded as processed files (Table 2.2). As such, the analysis pipeline for these libraries can be found within the corresponding paper and is distinct from the pipeline used to process in-house generated ATAC-seq data (see 2.5.2). Where '.wig' files were downloaded, I converted these to smaller files called '.bigwig' ('.bw') using the wigToBigWig binary available from the UCSC (Kent et al., 2010).

| Published Name | Thesis Name | Respository | Accession number | File | Description | Chapters Used |
|---|---|---|---|---|---|---|
| TS | "TS RL" | GEO | GSE94694 | TS_ATAC.wig.gz | Visual file | 3, 4 and 5 |
| | | | | TS_ATAC_DARs.bed.gz | Differential access peakset | 4 and 5 |
| | | | | TS_ATAC_total_peaks.bed.gz | Total accessible regions | 4 |
| d2diff | "-GF RL" | GEO | GSE94694 | d2diff_ATAC.wig.gz | Visual file | 4 and 5 |
| | | | | d2diff_ATAC_DARs.bed.gz | Differential access peakset | 4 and 5 |
| | | | | d2diff_ATAC_total_peaks.bed.gz | Total accessible regions | 4 |
| ZHBTC4 mESC | ZHBTC4 | GEO | GSE87819 | mESC_ZHBTC4_ATAC_UNT_MERGED .DANPOS.bgsub.smooth.bw | Visual file | 3 |
| ZHBTC4 24hDOX | ZHBTC4+DOX | GEO | GSE87819 | mESC_ZHBTC4_ATAC_24hrDOX_MERGED .DANPOS.bgsub.smooth.bw | Visual file | 3 |

**Table 2.2: Processed ATAC-seq libraries used throughout this thesis.**

All published ChIP-seq and RNA-seq libraries were downloaded as raw fastq files (Table 2.3) by Harshil Patel. RNA-seq analysis were processed using the same pipeline as outlined (see 2.5.4). ChIP-seq libraries were analysed in-line with the methods outlined (see 2.5.3) with minor alterations to accommodate for single-end reads.

| Description | Library Type | Respository | Accession number | Sample Identifier | Paired End / Single End | Chapters Used |
|---|---|---|---|---|---|---|
| ADACHI CDX2 | ChIP-seq | GEO | GSE51511 | GSM1246724 | Single End | 3 and 4 |
| ADACHI Input | ChIP-seq | GEO | GSE51511 | GSM1246726 | Single End | 3 and 4 |
| CHUONG CDX2 | ChIP-seq | GEO | GSE42207 | GSM1035392 | Single End | 3 |
| CHUONG Input | ChIP-seq | GEO | GSE42207 | GSM1035393 | Single End | 3 |
| H3K27ac | ChIP-seq | GEO | GSE42207 | GSM1035380 | Single End | 4 and 5 |
| | | | GSE42207 | GSM1035381 | Paired End | 4 and 5 |
| H3K4me3 | ChIP-seq | GEO | GSE42207 | GSM1035382 | Single End | 4 and 5 |
| H3K9me3 | ChIP-Seq | GEO | GSE42207 | GSM1035383 | Single End | 4 and 5 |
| | | | GSE42207 | GSM1035384 | Paired End | 4 and 5 |
| H3K4me1 | ChIP-Seq | GEO | GSE42207 | GSM1035385 | Single End | 4 and 5 |
| H3K27me3 | ChIP-Seq | GEO | GSE42207 | GSM1035386 | Single End | 4 and 5 |
| | | | GSE42207 | GSM1035387 | Paired End | 4 and 5 |
| Histone Input | ChIP-seq | GEO | GSE42207 | GSM1035393 | Single End | 4 and 5 |
| | | | GSE42207 | GSM1035394 | Paired End | 4 and 5 |
| TFAP2C in TSCs | ChIP-seq | ENA | PRJNA298763 | SRR2913350 | Single End | 4 and 5 |
| TFAP2C in DD1 | ChIP-seq | ENA | PRJNA298763 | SRR2913306 | Single End | 4 and 5 |
| TFAP2C Input | ChIP-seq | ENA | PRJNA298763 | SRR2913307 | Single End | 4 and 5 |
| TSC Rep 1 | RNA-seq | ENA | PRJNA298763 | SRR2895339 | Single End | 4 and 5 |
| TSC Rep 2 | RNA-seq | ENA | PRJNA298763 | SRR2895340 | Single End | 4 and 5 |
| TSC Rep 3 | RNA-seq | ENA | PRJNA298763 | SRR2895341 | Single End | 4 and 5 |
| TSC Rep 4 | RNA-seq | ENA | PRJNA298763 | SRR2895342 | Single End | 4 and 5 |
| TSC Rep 5 | RNA-seq | ENA | PRJNA298763 | SRR2895317 | Single End | 4 and 5 |
| DD1 Rep 1 | RNA-seq | ENA | PRJNA298763 | SRR2895318 | Single End | 4 and 5 |
| DD1 Rep 2 | RNA-seq | ENA | PRJNA298763 | SRR2895319 | Single End | 4 and 5 |
| DD1 Rep 3 | RNA-seq | ENA | PRJNA298763 | SRR2895320 | Single End | 4 and 5 |
| DD3 Rep 1 | RNA-seq | ENA | PRJNA298763 | SRR2895321 | Single End | 4 and 5 |
| DD3 Rep 2 | RNA-seq | ENA | PRJNA298763 | SRR2895322 | Single End | 4 and 5 |
| DD3 Rep 3 | RNA-seq | ENA | PRJNA298763 | SRR2895323 | Single End | 4 and 5 |

**Table 2.3: Summary of published ChIP-seq and RNA-seq libraries used in this thesis.** Datasets were downloaded as raw files and processed using the same pipelines as in-house generated data by Harshil Patel.

### 2.5.2 ATAC-seq analyses

Raw reads from each sample were adapter-trimmed using cutadapt (version 1.9.1) (Martin, 2011) with parameters "-a CTGTCTCTTATA -A CTGTCTCTTATA --minimum-length=25 –quality-cutoff=20". BWA (version 0.6.2) (Li and Durbin, 2010) with default parameters was used to perform genome-wide mapping of the adapter-trimmed reads to the mouse mm10 genome assembly downloaded from the UCSC (Karolchik et al., 2004). Read group addition, duplicate marking and insert size assessment was performed using the picard tools AddOrReplaceReadGroups, MarkDuplicates and CollectMultipleMetrics, respectively (version 2.1.1) (http://broadinstitute.github.io/picard/). Reads mapped to mitochondrial DNA were removed using the pairToBed command from BEDTools (version 2.26.0-foss-2016b) (Quinlan and Hall, 2010). Additional filtering was performed to only include uniquely mapped, properly-paired reads with insert size <= 1kb, and mismatches <= 1 in both

reads. SAMtools (version 1.3.1) (Li et al., 2009) was used for bam file sorting and indexing.

Three biological replicate libraries were generated per condition (Untransfected – UT, Scrambled mCherry negative – SCR NEG, Scrambled mCherry positive - SCR, *Cdx2* knockdown mCherry positive – KD). To generate sufficient numbers of reads, each biological replicate library was sequenced more than once. The filtered alignments from the technical replicates were merged for each biological replicate library at both the replicate level (per biological replicate) and sample level (per condition) using the picard MergeSamFiles command (http://broadinstitute.github.io/picard/). Duplicate marking and removal were reperformed on the merged alignments. BedGraph coverage tracks representing the accessibility signal per million mapped paired-reads were generated using BEDTools genomeCoverageBed with the parameters "-bg -pc -scale <SCALE_FACTOR>" for each biological replicate and for each condition (merged biological replicates). BedGraph files were converted to bigWig using the wigToBigWig binary available from the UCSC with the "-clip" parameter (Kent et al., 2010). Histogram of insert size frequency obtained from picard CollectMultipleMetrics and plotted using ggplot (version 2.2.1) within the R programming environment (version 3.3.1).

On both the biological replicates and condition replicates, regions of chromatin accessibility were identified genome-wide using MACS2 callpeak (version 2.1.1.20160309) (Zhang et al., 2008) with the parameters "--gsize=mm --keep-dup all –f BAMPE --nomodel --broad". A union set of intervals were obtained by merging the regions identified across all INX samples to generate a peak file containing all accessible regions called "merged all INX".

#### *2.5.2.1 Differential accessibility analysis*

Fragment-level BED files were derived from those created using the BEDTools (Quinlan and Hall, 2010) bamToBed command with the option "-bedpe". Differential chromatin accessibility sites between conditions were obtained using diffReps (version 1.55.4) (Shen et al., 2013) with the parameter "—frag 0". Differential sites that intersected with the union set of accessibility sites and had an FDR <= 0.01 and

fold-change >= 2 were kept for further analysis. The annotatePeaks.pl program from HOMER (version 4.8) (Heinz et al., 2010) was used to annotate the differential sites relative to mm10 RefSeq features downloaded from the UCSC on 19th February 2016.

#### *2.5.2.2 Footprinting Analysis*

Harshil Patel performed differential footprinting analysis on KD versus SCR and UT condition replicates using BaGFoot using default parameters (Baek et al., 2017). Bagplots (Rousseeuw and Ruts, 1997) were generated using a default factor of 3 (Baek et al., 2017).

### 2.5.3 ChIP-seq analyses

Raw reads from each sample were adapter-trimmed using cutadapt (version 1.9.1) (Martin, 2011) with parameters "-a CTGTCTCTTATA -A CTGTCTCTTATA --minimum-length=25 –quality-cutoff=20". BWA (version 0.6.2) (Li and Durbin, 2010) with default parameters was used to perform genome-wide mapping of the adapter-trimmed reads to the mouse mm10 genome. Duplicate marking was performed using the picard tool MarkDuplicates (version 2.1.1) (http://broadinstitute.github.io/picard/). Further filtering was performed to exclude read pairs that were duplicates, discordant, mapped to different chromosomes, ambiguously mapped, insert size >1kb, and had a mismatch >1 in any read.

Genome-wide peak calling was performed with MACS2 callpeak (version 2.1.1.20160309) (Zhang et al., 2008) with the parameters "--format BAMPE --gsize=hs --keep-dup all --cutoff-analysis --broad --broad 0.1". The annotatePeaks.pl program from HOMER (version 4.8) (Heinz et al., 2010) was used to annotate peaks relative to mm10 RefSeq features downloaded from the UCSC on 19th February 2016.

BedGraph coverage tracks representing the accessibility signal per million mapped paired-reads were generated using BEDTools (Quinlan and Hall, 2010) genomeCoverageBed with the parameters "-bg -pc -scale <SCALE_FACTOR>". BedGraph files were converted to bigWig using the wigToBigWig binary available from the UCSC with the "-clip" parameter (Kent et al., 2010).

### 2.5.4 RNA-seq analyses

Adapter trimming was performed with cutadapt (version 1.9.1) (Martin, 2011) with parameters "--minimum-length=25 --quality-cutoff=20 -a AGATCGGAAGAGC". The RSEM package (version 1.3.0) (Li and Dewey, 2011) in conjunction with the STAR alignment algorithm (version 2.5.2a) (Dobin et al., 2013) was used for the mapping and subsequent gene-level counting of the sequenced reads with respect to mm10 RefSeq genes downloaded from the UCSC Table Browser (Karolchik et al., 2004) on 19th February 2016. The parameters used were "--star-output-genome-bam --forward-prob 0".

#### *2.5.4.1 Differential expression analysis*

Differential expression analysis was performed with the DESeq2 package (version 1.12.3) (Love et al., 2014) within the R programming environment (version 3.3.1). An adjusted p-value of <= 0.05 was used as the significance threshold for the identification of differentially expressed genes. Gene set enrichment analysis for differentially expressed genes was performed by Gene Ontology Pathway and Biological processes using GeneGo MetaCore (https://portal.genego.com/).

### 2.5.5 Data visualisation and functional analyses

The Interactive Genomics Viewer (IGV) (version 2.4.8) (Robinson et al., 2011, Thorvaldsdottir et al., 2013) was used to visually explore ATAC-seq and ChIP-seq datasets and positioning of peaksets. From IGV, images of genomic regions were outputted as .svg files and cropped in Adobe Illustrator 22.1 (Adobe, USA).

#### *2.5.5.1 Motif analysis*

To find motifs within differential peaksets, motif calling analysis was performed using the findMotifsGenome.pl program in HOMER (version 4.8) (Heinz et al., 2010). Unless otherwise stated, this was performed using the inherent size of individual peaks using the parameter "-size given". All motif analysis was performed over "merged all INX" file described (see 2.5.2.1) as a background file set using "-bg" parameter. Visual motif files were generated using the ggseqlogo package (Wagih, 2017) in RStudio (version 1.0.136).

#### *2.5.5.2 Intersecting peaks*

To directly overlap peaksets, the mergePeaks.pl program was used from HOMER (version 4.8) (Heinz et al., 2010). Unless stated otherwise, peaks were directly overlapped with the parameter "-d given" and a file containing the number of overlapping peaks output using "-venn". Overlapping peaks were either visualised using Venn diagrams or UpSetR plots in RStudio (version 1.0.136). Venn diagrams were generated using the Vennerable package (version 1.0). UpSetR plots were generated using the UpSetR package (version 1.3.3) (Conway et al., 2017).

#### *2.5.5.3 GO analysis on ATAC-seq and ChIP-seq peaksets*

Gene ontology (GO) analysis on peaksets derived from ATAC-seq and ChIP-seq analyses was performed using Genomic Regions Enrichment of Annotations Tool (GREAT) (Version 3.0.0) (McLean et al., 2010). This tool assigns meaning to non-coding regions based on the genes closest to them. For this, the Association rule settings were set as the "Single nearest gene" to each peak. To provide contextual significance, all GREAT analysis was performed over the "merged all INX" file described (see 2.5.2.1) as a background file.

#### *2.5.5.4 DeepTools Visualisation*

All global visualisation of ATAC-seq and ChIP-seq libraries was performed using the tool suite deepTools (Ramirez et al., 2016).

To generate Pearson correlation heatmaps at (-bs) 5,000bp intervals for ChIP-seq libraries and (-bs) 10,000bp intervals for ATAC-seq libraries:

```
multiBamSummary bins --bamfiles <.bam> <> -o <result.npz> -bs #,
plotCorrelation --corData <result.npz> -c pearson -p heatmap -o
<Heatmap.pdf> --plotNumbers --colorMap YlGnBu -min #
```

To generate read coverage heatmaps displaying the average normalised read density at individual sites:

```
computeMatrix reference-point -S <.bw> <> -R <.bed> <> -a 800 -b 800 -bs
50 --samplesLabel <Label_1> <> -o <matrix.ma.gz> --referencePoint center,
```

```
plotHeatmap -m <matrix.ma.gz> -o <Heatmap.pdf> --refPointLabel center --kmeans # --outFileSortedRegions <ClusterOutput.bed> --colorMap Greens --heatmapHeight # --heatmapWidth #
```

To generate plotProfiles to show the average read density across all peaks in a given peakset, matrices were generated as described for read coverage heatmaps and subsequently used in the plotProfile tool with default settings.

# Chapter 3. Profiling Cdx2 expression dynamics and binding sites in trophoblast stem cells

E3.5 blastocysts and E6.5 extraembryonic ectoderm outgrowths can be cultured in the presence of critical growth factors to form stable trophoblast stem cell (TSC) lines (Tanaka et al., 1998). These lines express core markers of extraembryonic ectoderm, including *Cdx2*, and can be used to model the maintenance of trophoblast *in vitro*. Loss of *Cdx2* causes mouse embryos to die at approximately E4.5 (Strumpf et al., 2005). Null embryos have ectopic expression of pluripotency genes in the trophectoderm and defects in epithelial integrity, followed by the collapse of the blastocoel cavity (Strumpf et al., 2005). *Cdx2*-null embryos are unable to form any trophoblast outgrowth *in vitro* at E3.5 indicating *Cdx2* plays a role in TSC establishment and differentiation (Strumpf et al., 2005). Consistent with this suggestion, overexpression of *Cdx2* in mouse embryonic stem cells (mESCs) causes them to transdifferentiate into TSC-like cells (Niwa et al., 2005).

It is possible to overcome the immediate requirement for *Cdx2* for generating TSC-like cells from mESCs, either by repressing *Oct4* expression or by overexpressing other TSC marker genes that subsequently activate *Cdx2* expression (Niwa et al., 2005, Ralston et al., 2010, Kuckenberg et al., 2010, Rhee et al., 2014). However, TSC-like cells transdifferentiated in this way cannot self-renew in the absence of *Cdx2* expression. Instead, loss of *Cdx2* causes trophoblast stem cell markers such as *Eomes* to become down-regulated, and differentiation markers such as *Ascl2* to increase their levels of expression (Niwa et al., 2005). This suggests that *Cdx2* is a critical component of the gene regulatory network (GRN) that maintains trophoblast stemness and that this independent of its role in reinforcing the TSC fate.

Although much is known about its role in trophoblast lineage commitment, less is known about how *Cdx2* exerts its function in established TSCs. One approach, bearing in mind that CDX2 is a transcription factor, is to determine its genomic binding sites in established TSCs. Several CDX2 TSC ChIP-seq libraries have been published but these vary in quality, complexity and cell origin (Adachi et al., 2013, Chuong et al., 2013). In this chapter, I therefore characterise three embryo-derived

trophoblast stem cell lines that were subsequently used to generate CDX2 ChIP-seq libraries.

## 3.1 Characterisation of trophoblast stem cell lines

Three trophoblast stem cell (TSC) lines were characterised for their morphological appearance, stem cell marker gene expression, and rates of spontaneous differentiation. The INX and XYO cell lines were derived by Daniel Snell (The Francis Crick Institute), and the Rossant-GFP line was generated and kindly provided by Janet Rossant (Hospital for Sick Children, Canada).

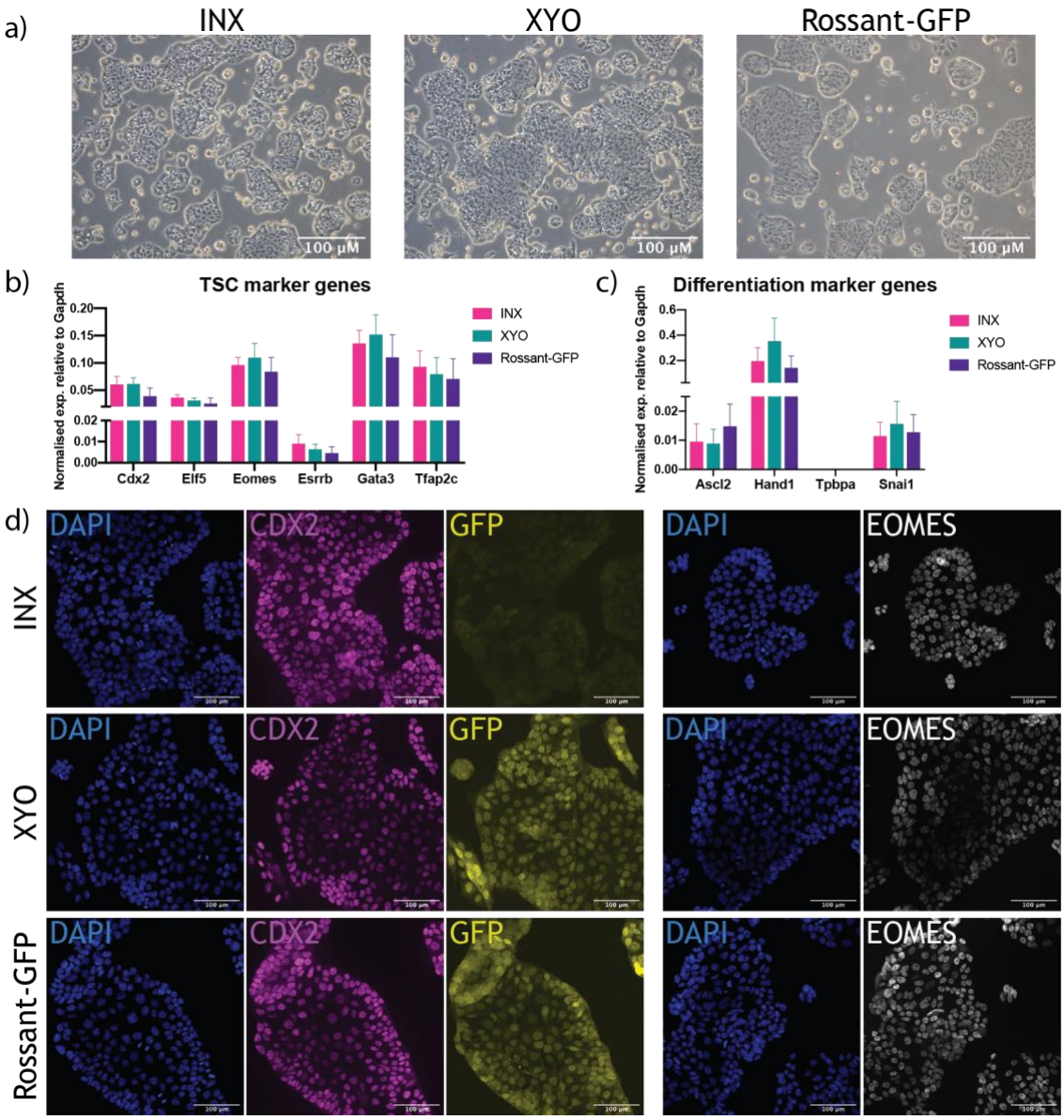



**Figure 3.1: Validation of TSC lines.** a) Bright-field images of the INX, XYO and Rossant-GFP TSC lines. b-c) qRT-PCR analysis of TSC and differentiation markers in all cell lines. Data plotted is gene expression normalised to *Gapdh*. Three biological replicates were used per cell line. Error bars display standard deviation (SD). d) Immunofluorescence for CDX2 and EOMES in three TSC lines.

All three cell lines share the morphological characteristics of TSCs, with cobblestone-like colonies and defined edges (Figure 3.1a). Expression of the TSC marker genes *Cdx2*, *Elf5*, *Eomes*, *Gata3* and *Tfap2c* was high in all three lines, although expression of the TSC marker gene *Esrrb* was low (Figure 3.1b). Consistent with this observation, expression of the differentiation genes *Ascl2*, *Tpbpa* and *Snai1* was low in each line, although expression of the trophoblast giant cell marker *Hand1* was high (Figure 3.1c). Expression of CDX2 and EOMES was confirmed by immunofluorescence, which suggested that CDX2 expression was higher at the edges of colonies (Figure 3.1d).

## 3.2 CDX2 is dynamically expressed in TSCs

I went on to examine CDX2 expression in colonies over time. Cells were plated at 1.05 x $10^4$ cells per $cm^2$ and sampled at 24-, 48- and 72-hours. Immunofluorescence showed that CDX2 protein is highly expressed at 24-hours in all three cell lines, but declined at 48- and 72-hours. CDX2 expression was higher in cells at the colony edge than in the centre at 48-hours in all three cell lines (Figure 3.2). This indicates that reduced CDX2 and a change in CDX2 distribution in colonies is a common feature of TSCs. There are three possible explanations for this decline in CDX2 expression: that it might depend on time of culture, cell density, or perhaps some sort of colony edge-effect. My experiments do not address which of these explanations is correct.

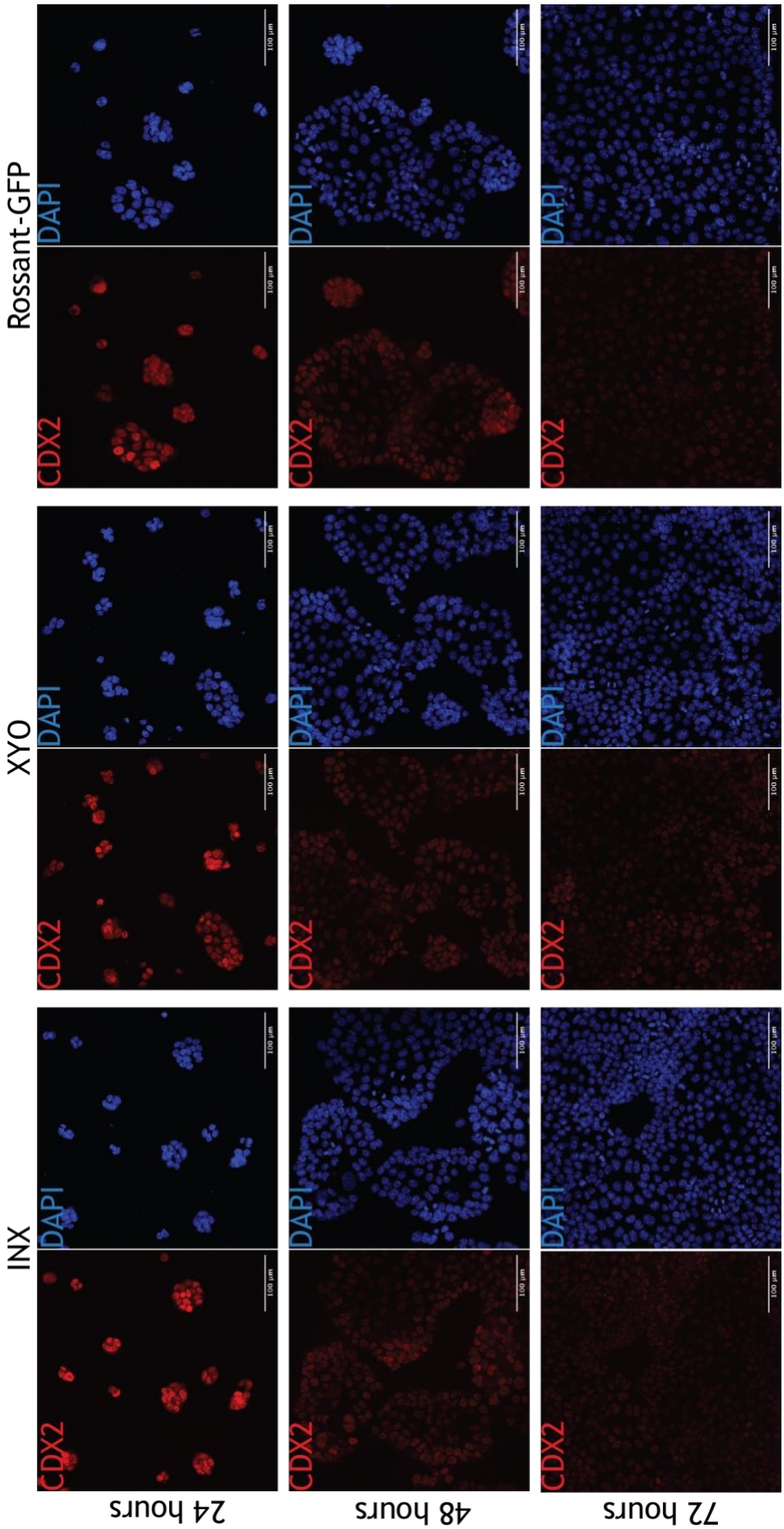



**Figure 3.2. CDX2 expression decreases with increased cell density.** Immunofluorescence staining for CDX2 in INX, XYO and Rossant-GFP cell lines at 24-, 48- and 72-hours post-cell plating

To further investigate the decline in CDX2 expression over time, I performed flow cytometry. CDX2 expression in single cells was determined using the gating strategy indicated in Figure 3.3a. As shown in Figure 3.3b, CDX2 expression levels were higher in cells at G2/M than in cells in G1. Some proteins, including C-MYC, are regulated by the cell cycle and increase from G1 to S to G2/M (Gookin et al., 2017), and although I have not assayed for a direct relationship between CDX2 and the cell cycle, it is possible that there is one (Figure 3.3b). Beyond this, I confirmed that CDX2 expression decreased significantly in all cell lines from 24- to 72-hours post-plating (Figure 3.3c-d). As with CDX2 protein, *Cdx2* transcripts show a decreased expression trend at each time point in all TSC lines (Figure 3.3e). When the TSC cell lines replicates are combined, *Cdx2* transcript decreased significantly between 24-hours and 72-hours post plating (Figure 3.3f).

To ask whether increased confluency resulted in cell cycle arrest, cells were classified into G1, S and G2/M phases using the 'biology tool' in FlowJo (Figure 3.3g). The proportion of cells in each cell cycle stage did not change between time points, suggesting that the cell cycle is not altered by time/cell density/edge (Figure 3.3h). Thus, the decrease in CDX2 expression between 24-hours and 72-hours post-plating is not due to cell cycle arrest in G1.

I next asked whether the decline in CDX2 expression is peculiar to a specific stage of the cell cycle or whether it occurs across all stages. To this end, I calculated the fold change in mean CDX2 expression at each cell cycle stage over the mean expression in the total sample (Figure 3.3i). This revealed that neither cell cycle arrest nor a decline of CDX2 at a given stage of the cell cycle causes the decline in CDX2 expression over time.

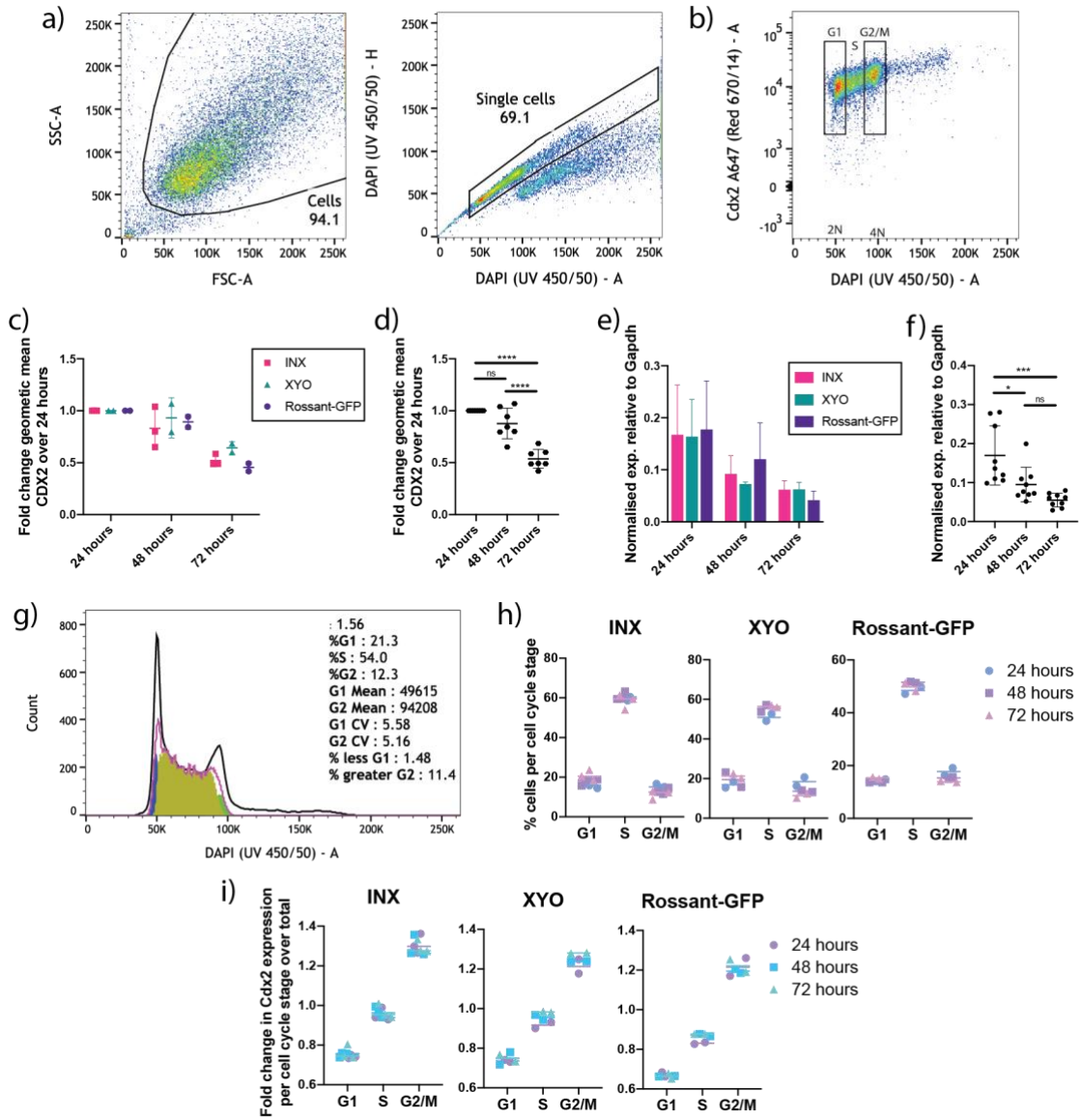



**Figure 3.3: Downregulation of Cdx2 at the protein and transcript level with increased cell density.** Dot plots showing gating strategy used for flow cytometry analyses. SSC-A and FSC-A were used to identify the intact cell population. Singlet discrimination was completed using DAPI-A and DAPI-H. Nuclei size (DAPI-H) is proportional to its area (DAPI-A). Thus, if the area doubles relative to height, cells are excluded from the analysis. b) Example dot plot showing CDX2 expression throughout the cell cycle (indicated by DAPI). Diploid cells (2N) sit at 50K in G1. The DNA is replicated during S phase, resulting in 4N cells (100K) at G2/M. The concentration of antibody was adjusted based on cell number so that the ratio between cell to antibody was the same between time points and biological replicates. c-d) Fold change in geometric mean CDX2 expression normalised to the geometric mean at 24 hours for individual cell lines (c) and in all cell lines combined (d) (n=3 for INX, n=2 for XYO, n=2 for Rossant-GFP). Significance determined by one-way ANOVA statistical test: **** = <0.0001. Error bars display SD. e-f) Normalised relative *Cdx2* expression in INX, YXO and Rossant-GFP at 24- 48- and 72-hours post-plating for individual cell lines (e) and in all cell lines combined (f) (n=3 per cell line). Error bars display SD. Statistical significance was determined using one-way ANOVA: * = 0.0146, *** = 0.0003. g) Example cell cycle histograms performed using the 'cell cycle biology tool' in FlowJo. h) Proportion of cells in each cell cycle stage. Error bars display SD. i) Fold change of average geometric mean for each cell cycle stage normalised to the total geometric mean CDX2 expression. Error bars display SD.

qRT-PCR was performed for other TSC markers to determine whether their expression also decreased over time. Like *Cdx2*, *Eomes* and *Gata3* expression decreased from 24- to 72-hours after plating (Figure 3.4a-b). This decline was confirmed for GATA3 protein (Figure 3.4f). Levels of *Esrrb* also showed a small decrease, but this was not significant (Figure 3.4c). In contrast, *Elf5* and *Tfap2c* gene expression remained the same at all timepoints (Figure 3.4d-e). Together, these observations reveal that some, but not all, TSC marker genes expression decreased in TSC cultures over time.

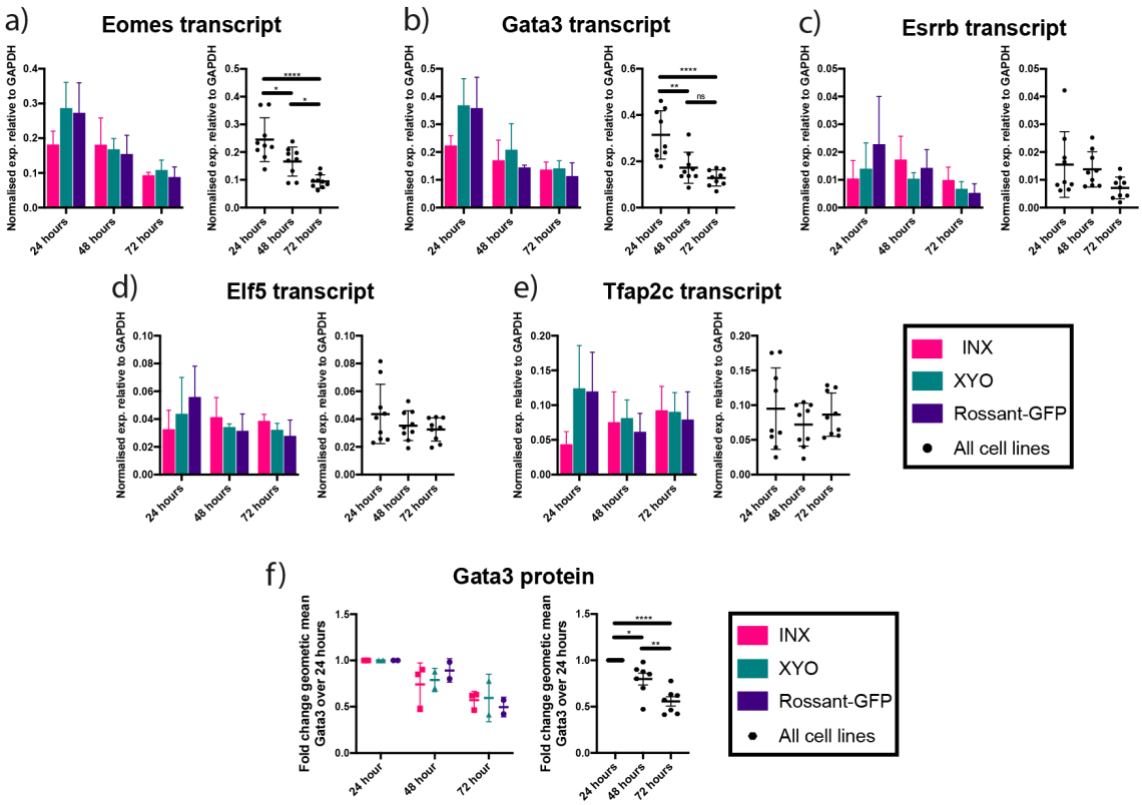



**Figure 3.4: TSC markers have different expression dynamics relative to cell density.** a-e) qRT-PCR analysis of TSC marker genes at 24-, 48- and 72-hours post-cell plating. Data are normalised to *Gapdh* individual cell line (left) and for combined TSCs (right) (n=3 biological replicates per cell line). Significance determined by one-way ANOVA statistical test: * = <0.0332, ** = <0.0021, *** = <0.0002, **** = <0.0001. Error bars display SD in both combined and separate graphs. f) Individual (left) and combined (right) fold change in geometric mean GATA3 expression relative to 24 hours (INX n=3, XYO n=2, Rossant-GFP n=2). Significance determined by one-way ANOVA statistical test: * = 0.0175, ** = 0.005, **** = <0.0001. Error bars display SD.

## 3.3 Determining CDX2 binding sites in TSCs

In an effort to decipher the role *Cdx2* plays within the established trophoblast lineage, I determined its binding targets. Several CDX2 ChIP-seq datasets have been published for TSCs (Table 3.1). As shown in Table 3.1, I generated new CDX2 ChIP-seq libraries for both the INX and XYO cell lines. For this, I re-validated a previously published ChIP-grade anti-CDX2 antibody (A300-691A, Bethyl laboratories) in accordance with ENCODE guidelines (Landt et al., 2012) using immunofluorescence (Figure 3.2) and immunoprecipitation/western blot (data not shown).

In Chapter 3.2, I showed that CDX2 expression declines in TSCs the longer they are in culture. TSCs were therefore collected at approximately 60-70% confluency as a compromise between cell number and CDX2 expression. Two independent ChIP-seq libraries were generated for the INX and XYO cell lines on two separate occasions (yellow, Table 3.1). Two published datasets were downloaded for incorporation into analyses: a Rossant-derived TSC line dataset (Chuong et al., 2013) and a dataset from a transdifferentiated TSC line generated using the tet-inducible *Oct3/4*-knockout cell line ZHBTc4 (Adachi et al., 2013) (blue, Table 3.1).

| Name | Type | Antibody | GEO Series | Total Input Pairs | % Mapped | %Duplication | Filtered Total Reads | NarrowPeaks |
|---|---|---|---|---|---|---|---|---|
| INX_R1 | Cdx2 ChIP | A300-691A, Bethyl | - | 50025598 | 96.41% | 8.50% | 33249212 (66.77%) | 31895 |
| INX_R2 | Cdx2 ChIP | A300-691A, Bethyl | - | 41850024 | 96.16% | 7.59% | 28127039 (67.53%) | 26220 |
| INX_INPUT | Input | - | - | 56984844 | 96.73% | 5.19% | 39592108 (69.96%) | - |
| XYO_R1 | Cdx2 ChIP | A300-691A, Bethyl | - | 25695989 | 96.35% | 7.17% | 17510914 (68.51%) | 2891 |
| XYO_R2 | Cdx2 ChIP | A300-691A, Bethyl | - | 69423294 | 96.77% | 8.62% | 43974816 (63.64%) | 23886 |
| XYO_INPUT | Input | - | - | 55532306 | 96.72% | 5.98% | 33055761 (63.90%) | - |
| ADACHI | Cdx2 ChIP | CDX2-88, BioGenex | GSE51511 | 25776163 | 97.41% | 9.19% | 19558086 (76.24%) | 103992 |
| ADACHI_INPUT | Input | - | GSE51511 | 26451066 | 97.14% | 9.63% | 19761785 (74.77%) | - |
| CHUONG | Cdx2 ChIP | A300-691A, Bethyl | GSE42207 | 16534562 | 85.61% | 61.20% | 3272500 (20.07%) | 1171 |
| CHUONG_INPUT | Input | - | GSE42207 | 69336966 | 92.54% | 35.86% | 29469243 (44.06%) | - |

**Table 3.1: CDX2 ChIP-seq dataset alignment statistics.** Summary of alignment figures during subsequent filtering steps for in-house generate (yellow) and published (blue) Cdx2 ChIP-seq libraries analysed.

All libraries were aligned to the mm10 genome, filtered and normalised using the approach outlined in Materials and Methods (Chapter 2.5.3). Table 3.1 summarises alignment efficacy and the number of peaks called for each ChIP library over the relevant input library using MACS2. This analysis revealed that the CHUONG CDX2 ChIP-seq dataset is of particularly low complexity, with >61% of total starting reads attributed to PCR duplicates. After quality filtering and duplicate removal, only 20% of the total reads remained in this library. With so few reads remaining, read coverage

was poor, leading to poor clustering with the other datasets, absence of called and visible peaks on IGV tracks, and minimal read density enrichment across total called peaks (Figure 3.5a-c). This library was therefore omitted from further analyses.

INX and XYO TSC libraries correlated with well one another. However, correlation is weaker for XYO_R1, perhaps because this library contained almost half the number of reads as other in-house generated libraries (Figure 3.5a, Table 3.1). The effect of the reduced number of reads in XYO_R1 was that it produced a lower signal-to-noise ratio. This lower ratio does not change the position of CDX2 binding sites between libraries, but does change how many normalised reads are found in each peak within the XYO_R1 library (Figure 3.5b-c). Overall, the INX and XYO CDX2 ChIP-seq libraries confirm that CDX2 binds at the same binding sites between cell lines.

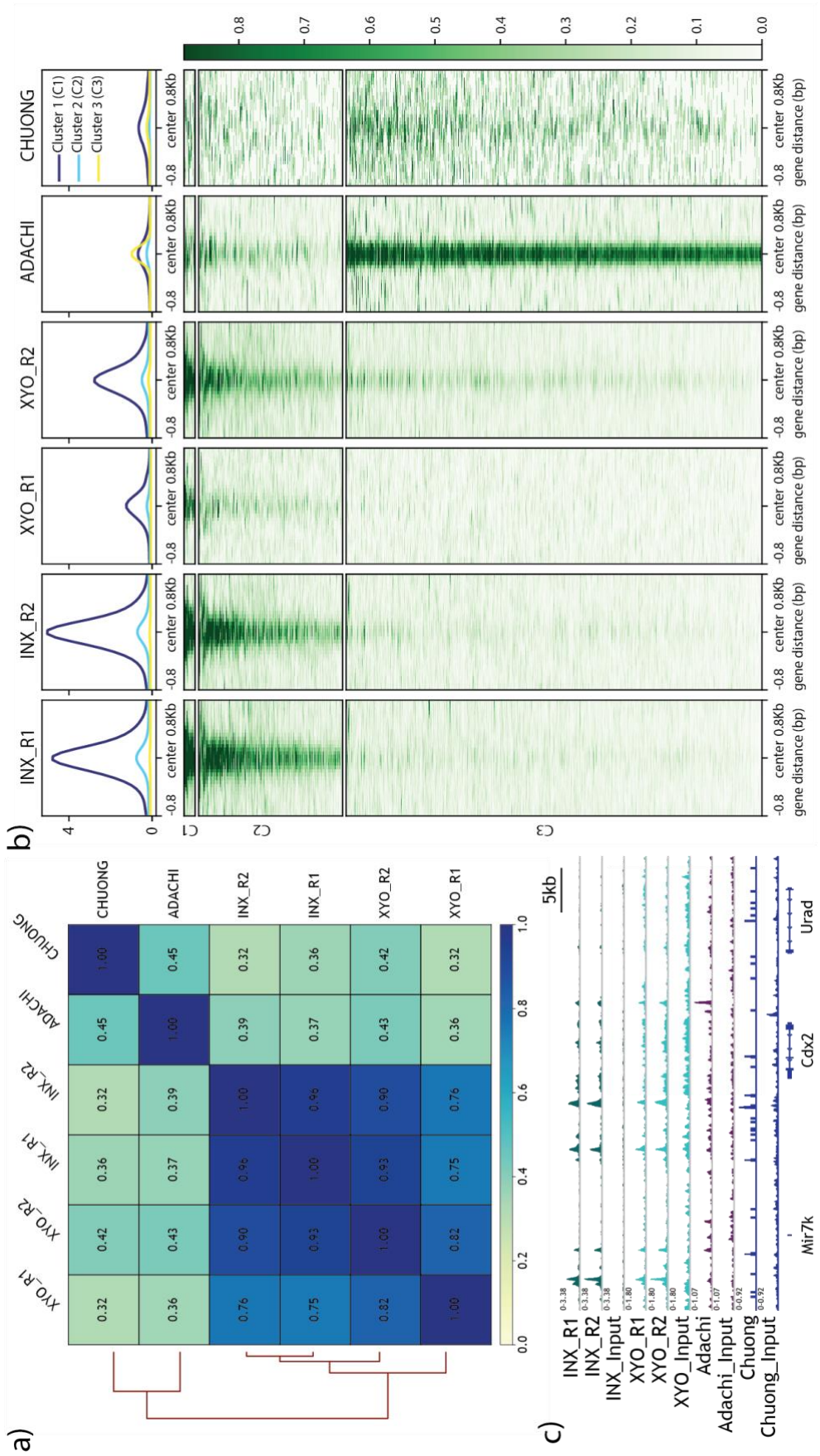



**Figure 3.5: Comparison of CDX2 ChIP-seq libraries generated in established TSCs.** a) A correlation heatmap showing Pearson correlation coefficients calculated using read density in 5000bp intervals across the entire genome. b) Heatmap for normalised read density at CDX2 binding sites. All peaksets determined for INX, XYO and ADACHI libraries relative to their respective inputs were merged to make a total CDX2 binding site peakset used to generate this heatmap. c) IGV (Integrative genomics viewer) track showing CDX2 binding around the *Cdx2* gene for all six ChIP-seq libraries analysed.

The ADACHI library correlates poorly with the INX and XYO libraries, instead clustering separately with the discarded CHUONG library (Figure 3.5a). As with the XYO_R1 library, this library contains a reduced number of total reads. In contrast to the XYO_R1 library, over 100,000 CDX2 binding peaks were called in the ADACHI library indicating that it has a good signal-to-noise ratio (Table 3.1), although about 89% of ADACHI peaks are unique (Figure 3.6a). These differences in CDX2 binding sites between the ADACHI and INX/XYO libraries were confirmed by the different position and height of peaks between their tracks (Figure 3.5c).

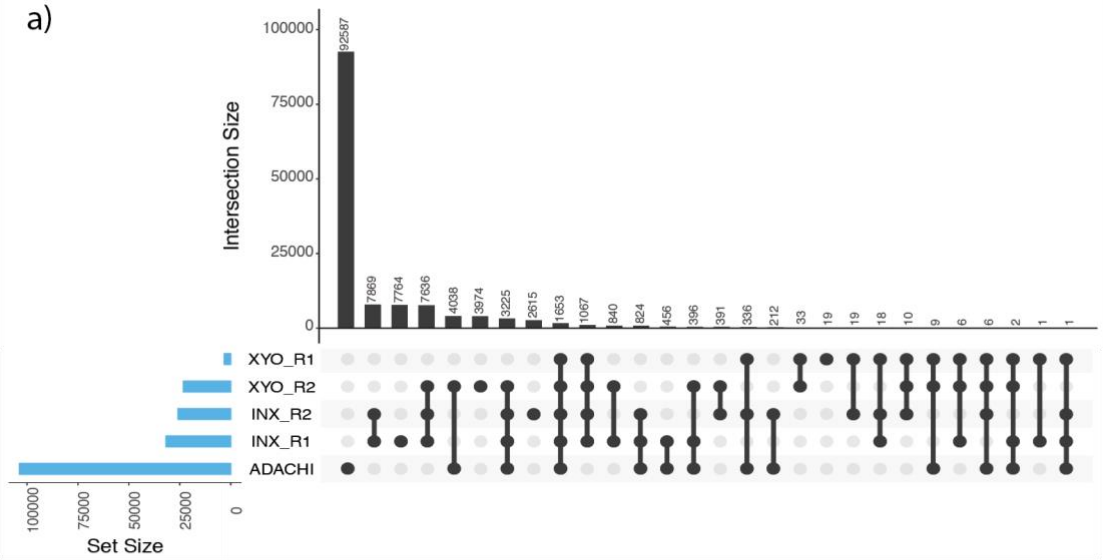
Therefore, I next performed motif analysis on the ADACHI, INX and XYO libraries to confirm that CDX2 is binding its consensus motif in all of these libraries (Heinz et al., 2010). CDX2 contains a helix-turn-helix (HTH) homeodomain which recognises a consensus motif sequence that is similar to, and shared with, other HTH homeodomain family members (Burglin and Affolter, 2016, Catron et al., 1993, Mathelier et al., 2016). Whilst 90.6% of the ADACHI peaks contained the CDX2 consensus motif, fewer than half of INX peaks contained this motif (Figure 3.6b).

To examine the relationships between all the peaks called across these five libraries, K-means clustering was performed for read density heatmaps. This best separated the data into three clusters: C1, C2 and C3. In contrast to the high enrichment of the INX and XYO libraries in clusters 1 and 2, the ADACHI library showed low read density enrichment in clusters 1 and 2, but was high in cluster 3 (Figure 3.5b). Therefore, Cluster 1 (C1) contained ‘universal’ peaks, whilst cluster 2 (C2) and cluster 3 (C3) contained mainly INX/XYO (Embryo-derived TSC-specific peaks) and ADACHI (Transdifferentiated TSC-like-specific peaks), respectively (Figure 3.5b).

There are two possible explanations for these differences in CDX2 binding and motif enrichment between the ADACHI and the INX/XYO libraries. Firstly, these libraries are generated from TSCs cells that were derived using different methods and the differences may, therefore, be biological in nature. Whilst the INX and XYO TSC lines are derived from embryo outgrowths, ADACHI cells were TSC-like cells that transdifferentiated from mESCs by repressing *Oct4* expression. Although these transdifferentiated cells were said to be stable cell lines, how closely they resemble

*bona fide* TSCs is unknown (Adachi et al., 2013). Alternatively, differences in CDX2 binding and motif enrichment may be an artefact of the different antibodies used to make libraries (Table 3.1).

a)




b)

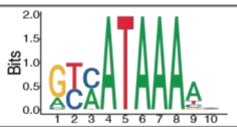
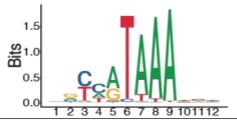
| Motif | Percentage of targets | | | | |
|---|---|---|---|---|---|
| | INX_R1 | INX_R2 | XYO_R1 | XYO_R2 | ADACHI |
| Cdx2  | 42.35% | 45.19% | 63.96% | 65.06% | 90.62% |
| Cdx4  | 49.02% | 52.30% | 70.29% | 69.23% | 86.24% |

**Figure 3.6: Embryo-derived and transdifferentiated CDX2 binding sites are different in location and motif enrichment.** a) Plot generated by UpSetR showing the number of overlapping peaks between individual narrowPeak peak files. b) Percentage of peaks per ChIP-seq library containing the consensus *Cdx2* and *Cdx4* motifs called using HOMER.

## 3.4 CDX2 has different roles in trophoblast stem cells depending on their origin

The ability of a transcription factor to bind a genomic region depends on many factors. Transcription factors bind to consensus sequence motifs (Khan et al., 2018). Whilst there many consensus sequences across the genome, only a small proportion are typically occupied *in vivo* at any given time (Carr and Biggin, 1999, Iyer et al., 2001, Joseph et al., 2010). Factors that influence binding include the accessibility of given chromatin regions, the histone modifications present and the occupancy of other factors in the vicinity (Pique-Regi et al., 2011, Kumar and Bucher, 2016, Jolma et al., 2015). These factors can be assayed using next-generation sequencing techniques; whilst ChIP-seq determines the distribution of a transcription factor or histone mark across the genome, FAIRE-Seq, DNase-Seq and ATAC-seq assess whether or not chromatin is accessible (Solomon et al., 1988, Giresi et al., 2007, Crawford et al., 2006, Buenrostro et al., 2013). From these analyses, it is possible to work out which binding sites are likely to be used within a given cell type.

So-called pioneer factors are exceptions to the rules that govern transcription factor binding. These pioneer factors are able to bind to condensed chromatin regions and cause these regions to open. Their binding precedes that of other factors and can be either passive or active: while passive binding recruits other factors that open chromatin regions, active binding opens chromatin itself and allow other factors to subsequently bind (Cirillo et al., 2002, Zaret and Carroll, 2011).

There is conflicting evidence on whether CDX2 is a pioneer factor (McPherson et al., 1993, Mahony et al., 2014, Kumar et al., 2019, Verzi et al., 2013, Rhee et al., 2017). If *Cdx2* is overexpressed in different cell types, CDX2 binds to its consensus motif in chromatin that is already accessible, which results in distinct CDX2 binding between conditions (Mahony et al., 2014, Rhee et al., 2017). This suggests that CDX2 has minimal pioneer capacity when overexpressed in non-CDX2 expressing cell types.

TSCs are an established, robust *in vitro* cell line (Tanaka., 1998). Therefore, I next asked whether CDX2 binding sites in INX, XYO and ADACHI binding sites are accessible in TSCs. For this, I generated a read density heatmap at the previously

determined CDX2 binding site clusters (C1, C2 and C3, Figure 3.5b) in two wild-type TSC libraries: the untransfected INX ATAC-seq library generated for this thesis (data further discussed in Chapter 4) and a published TSC ATAC-seq library ('TS RL') (Nelson et al., 2017) (Figure 3.7a).

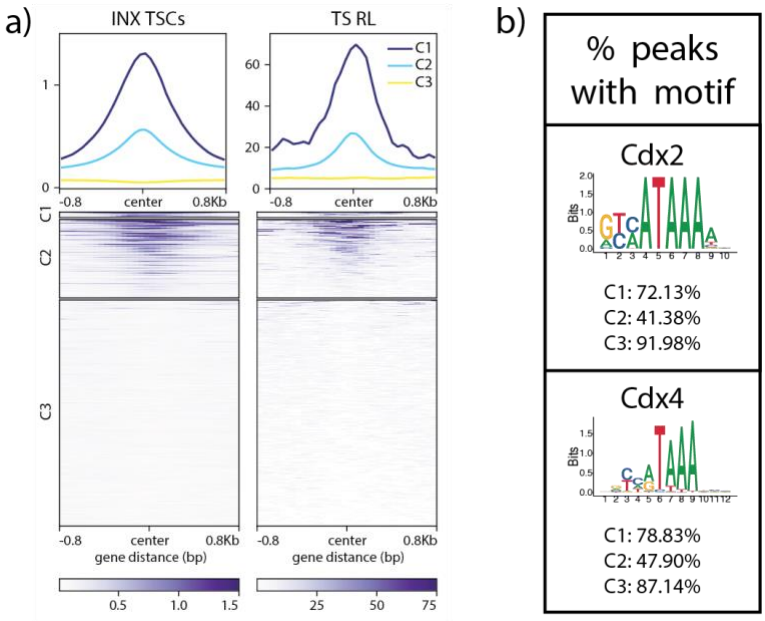



**Figure 3.7: Deciphering the accessibility of distinct CDX2 binding sites.** a) Heatmap of ATAC-seq read density in Untransfected (UT) INX TSCs and published TSCs called 'TS RL' at clusters from Figure 3.5b. The methods used to process and normalise libraries can change the scale used to visualise libraries, although this has no bearing on library quality, nor on enrichment. The 'TS RL' ATAC-seq data was downloaded as aligned visualisation files (.wig) and so it is on a different scale to the INX data (Nelson et al., 2017). b) Percentage of peaks per cluster containing the consensus *Cdx2* and *Cdx4* motifs determined using HOMER. C1 – cluster 1, C2 – cluster 2, C3 – cluster 3.

Cluster C1 sites were bound in all CDX2 ChIP-seq libraries and were considered to be universal CDX2 binding sites. In contrast, sites in cluster C2 are bound by CDX2 in both INX and XYO TSCs, but not ADACHI cells and so represent embryo-derived TSC-specific CDX2 binding sites. As C3 binding sites are only bound in ADACHI cells, they represent transdifferentiated TSC-like cell-specific peaks (Figure 3.5b).

Whilst most peaks in clusters C1 and C2 are accessible in INX and/or 'TS RL', all cluster C3 sites are inaccessible in both (Figure 3.7a). Although cluster C3 sites are highly enriched for the CDX2 consensus motif, it is unlikely that CDX2 binds 99,177 inaccessible regions of chromatin in *bona fide* TSCs (Figure 3.7b). Instead, C3 may represent accessible regions in mESCs that CDX2 binds when expressed, suggesting that the differences in CDX2 binding are biological.

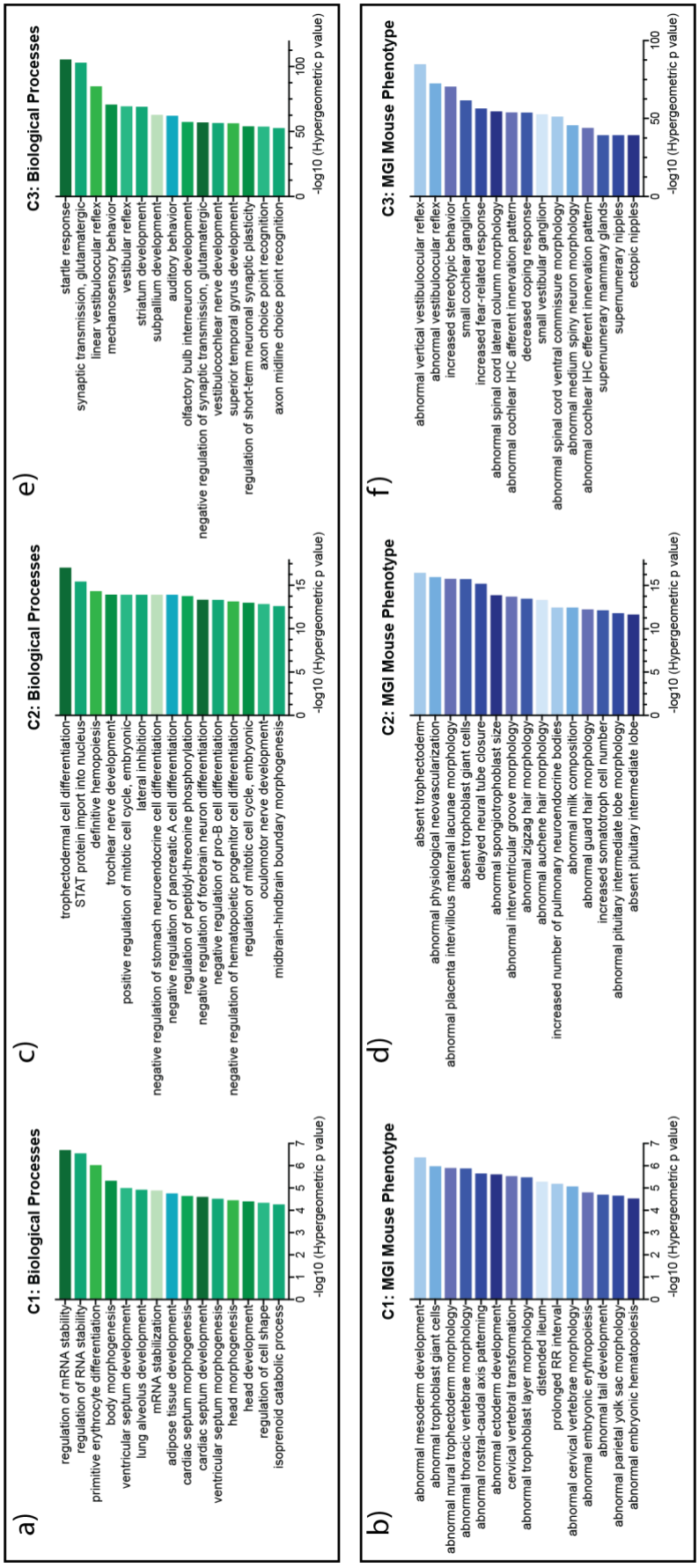



**Figure 3.8: GO analysis of TSC binding site clusters defined by derivation method.** a) Bar charts showing -log10 of the hypergeometric p-value for biological GO in clusters C1, C2 and C3 using the single nearest gene association rule setting by GREAT. b) Bar charts showing -log10 of the hypergeometric p-value for MGI Mouse Phenotype using GREAT.

To assess potential functional differences between peaks contained in each cluster, gene ontology (GO) analysis was performed using GREAT (McLean et al., 2010). In setting the parameters for this, I found the single nearest gene for each peak within each cluster. The enrichment was performed over a background file containing all regions deemed accessible in TSCs ATAC-seq data (Chapter 2.5.2).

Cluster 1 peaks are enriched around genes associated with abnormalities in the trophoblast lineage, including abnormal trophoblast giant cells (MP:0005033), abnormal mural trophectoderm morphology (MP:0012057) and abnormal trophoblast layer morphology (MP:0005031) (Figure 3.8a-b). Similarly, cluster 2 sites are enriched around genes related to absent trophoblast lineages, including absent trophectoderm (MP:0012102) and absent trophoblast giant cells (MP:0001714), as well as various abnormal trophoblast lineage morphology. Further, the most enriched biological process term found in cluster 2 was trophectodermal cell differentiation (GO:0001829), reflecting a potentially specific role of these sites within established TSCs (Figure 3.8c-d). In contrast, the top 100 terms discovered in cluster 3 showed no enrichment for trophoblast-associated biological processes (Figure 3.8e), MGI mouse phenotypes (Figure 3.8f), nor trophoblast related lineages in the top 100 MGI expression terms (data not shown). As such, this analysis provided no distinct biological function for transdifferentiated TSC-like cell-specific (ADACHI) CDX2 binding sites.

One potential interpretation of these differential CDX2 binding sites is that CDX2 has different targets before and after trophoblast commitment, represented by embryo outgrowth-specific (C2) and transdifferentiated-specific (C3) sites, respectively. As *Cdx2* has a reciprocal, repressive relationship with *Oct4* (encoded by the *Pou5f1* gene) (Niwa et al., 2005) and *Nanog* (Chen et al., 2009b) during trophectoderm establishment, I examined the flanking regions of these gene loci for differences in CDX2 binding (Figure 3.9). Near the transcription start site (TSS) of both *Pou5f1* and *Nanog*, there is a highly enriched CDX2 binding site in INX and XYO_R2 samples, with other peaks present in the vicinity. Corresponding peaks are present at these sites in TSC ATAC-seq libraries, showing that these sites are accessible. In contrast, CDX2 binding in the ADACHI (transdifferentiated TSC-like cells) was distinct from INX and XYO and those regions inaccessible in TSCs. Therefore, CDX2 remains

bound around pluripotency genes even in established TSCs, but these sites are not all equivalent to those found in transdifferentiated cells generated by *Oct4* removal.

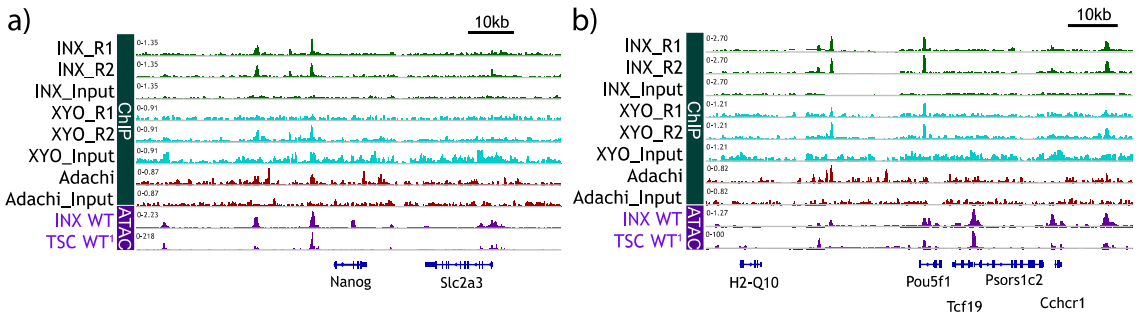



**Figure 3.9: CDX2 binding around pluripotency genes.** a) Integrative Genomics Viewer (IGV) tracks at pluripotency gene *Nanog*. B) IGV track for the Oct4 encoding gene *Pou5f1* locus.

## 3.5 Cdx2 may have different roles pre- and post-trophoblast establishment

*Pou5f1* and *Nanog* may be repressed by different CDX2 binding sites during and after trophoblast establishment. Alternatively, differences in CDX2 binding between embryo-derived TSCs (INX/XYO) and TSC-like cells transdifferentiated from mESCs (ADACHI) may be due to difference in global chromatin accessibility between TSCs and mESCs, respectively.

ADACHI CDX2 ChIP-seq libraries were generated from transdifferentiated ZHBTc4 mESCs. To determine if CDX2 binding in transdifferentiated TSC-like cells are accessible in ZHBTc4 mESCs, read density heatmaps were generated for published wild-type mESCs and 24-hours post-*Oct4* knockout mESCs at C1, C2 and C3 sites (Figure 3.10a) (King and Klose, 2017). CDX2 binding sites in the ADACHI-specific cluster (C3) are not bound in these mESCs, nor one day after *Oct4*-knockout.

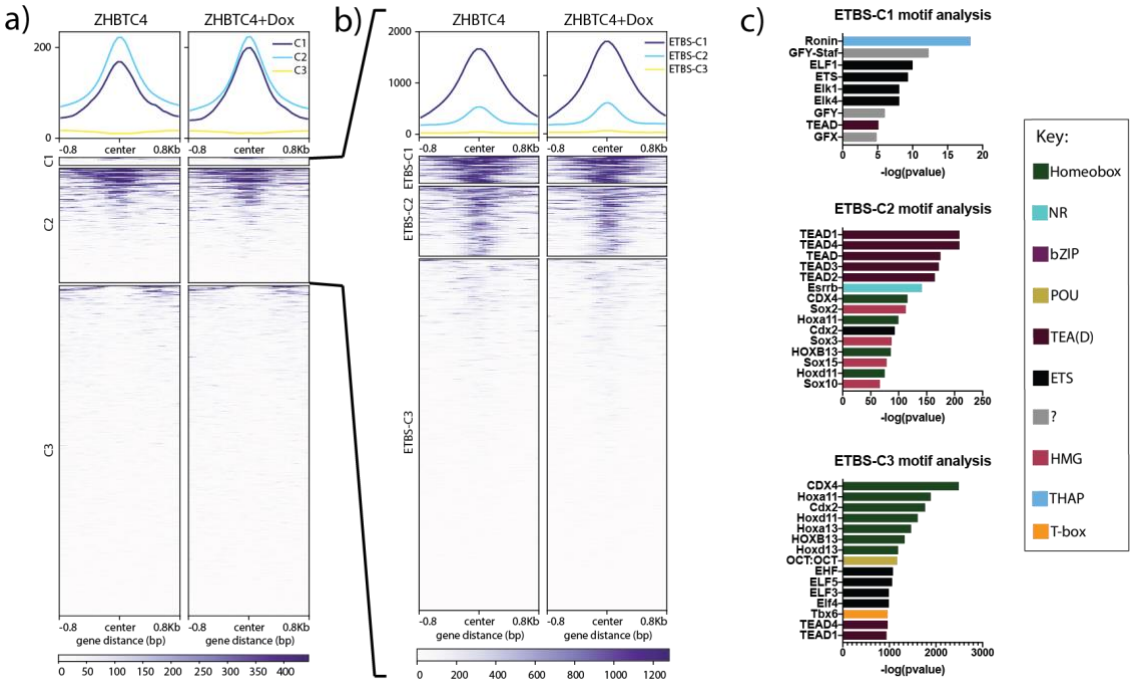



**Figure 3.10: Accessibility of CDX2 binding sites in mESCs.** a) ATAC-seq read density heatmap at C1, C2 and C3 clusters (defined in Figure 3.5b) for ZHBTc4 mESCs and ZHBTc4 mESCs 24-hours post doxycycline treatment (ZHBTc4 +Dox) to delete Oct4. b) Normalised read density heatmap to show accessibility at peaks contained in C1 and C2, collectively called established trophoblast CDX2 binding sites (ETBS), in ZHBTc4 and ZHBTc4-Dox mESCs. K-means clustering was performed to generate three new clusters: ETBS-C1, ETBS-C2 and ETBS-C3. c) Bar charts to show -log(p value) of highest enriched motifs discovered in each ETBS cluster as determined by HOMER analysis using given peak sizes.

Conversely, those CDX2 binding sites in embryo-derived TSCs, clusters C1 and C2, show high read density for mESC ATAC-seq data at some peaks, but negligible at others. At peaks contained in C1, overall read density increased after *Oct4*-knockout (summary plot, Figure 3.10a). Therefore, C1 sites may be those that CDX2 binds to in order to reinforce trophoblast fate.

Does the accessibility of CDX2 binding sites in mESCs reveal information about their function? If sites are crucial for reinforcing the trophoblast fate, CDX2 would likely be recruited by its consensus motif and other known upstream co-factors. In contrast, CDX2 may bind opportunistically at regions of accessible chromatin in the absence of its consensus motif. As such, C1 and C2 peaks were combined to make an embryo-derived TSC CDX2 binding site (ETBS) peakset. K-means clustering was performed to separate ETBS sites into 3 new clusters based on their accessibility in mESCs: highly accessible, accessible and inaccessible sites (Figure 3.10b).

Motif analysis revealed that highly accessible chromatin regions in mESCs (ETBS-C1) show no enrichment for homeobox motifs (Figure 3.10c). Sites with lower-level accessibility in mESCs, ETBS-C2, show enrichment for the CDX2 consensus motif (contained in 32.2% of peaks) and other homeodomain motifs. However, the five most significantly enriched motifs were TEA(D) family motifs (Figure 3.10c). TEAD4 is upstream of *Cdx2* in trophoblast specification *in vivo* and directly interacts with CDX2 in established TSCs (Nishioka et al., 2009, Rayon et al., 2014, Latos et al., 2015a). Motifs for other core TSC marker genes, including *Esrrb* and *Sox2*, were also enriched in this C2 cluster. The enrichment of these motif families indicates that TEAD4, ESRRB and SOX2 may be CDX2 co-factors in TSCs. As such, these peaks may represent sites crucial for reinforcing commitment to the trophoblast fate. In contrast, the top 7 enriched motifs found in inaccessible sites in mESCs (ETBS-C3) are homeobox motifs (Figure 3.10c). This cluster may therefore represent CDX2 binding sites in the established trophoblast fate. The co-enrichment of *Elf5* motifs, another key TSC marker, is also consistent with this function (Latos et al., 2015b). The enrichment of the POU motif Oct:Oct in the cluster C3 may suggest a small role for these sites in establishing the fate as a result of CDX2 antagonising OCT4 binding at these sites (Figure 3.10c).

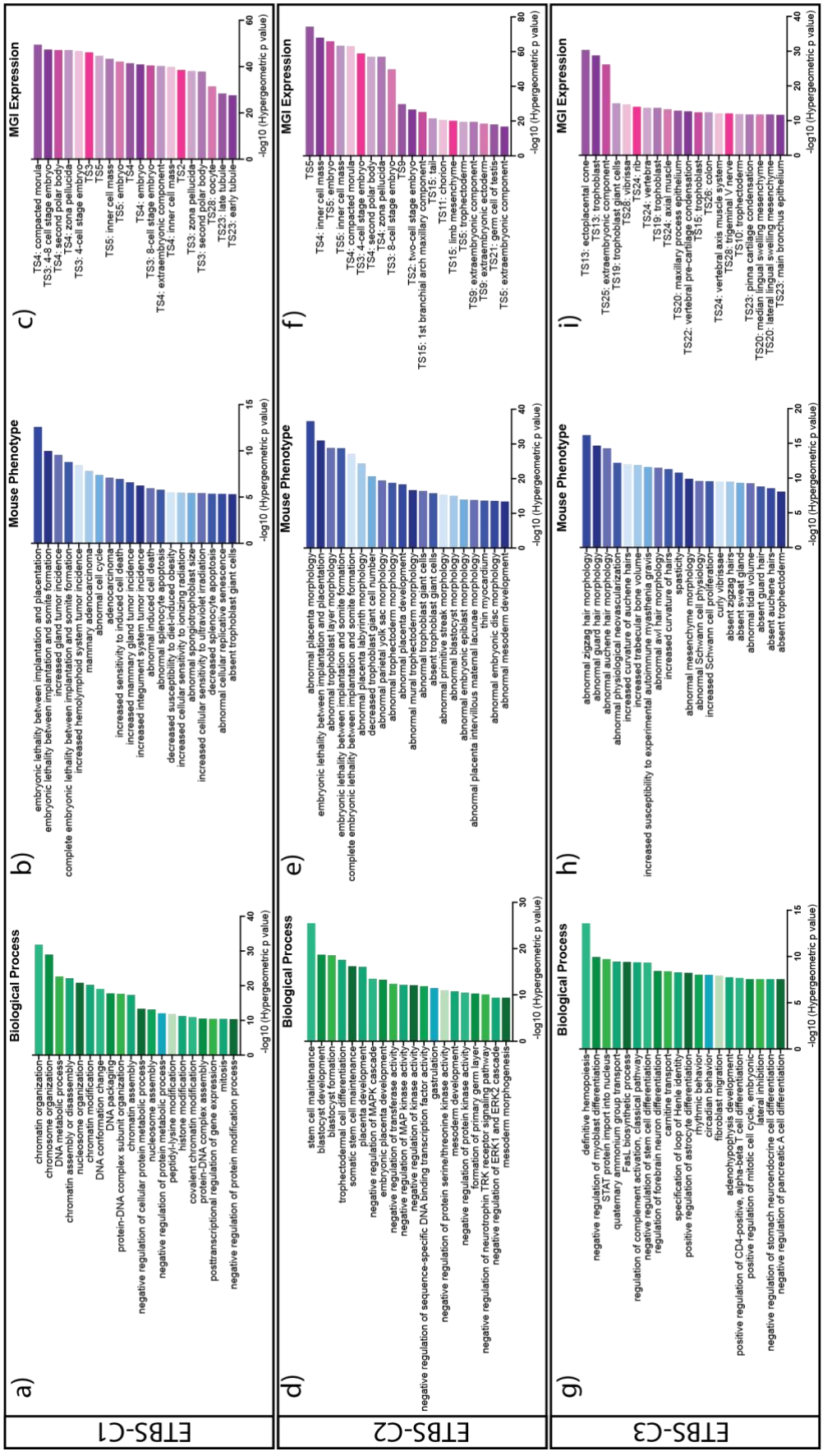



**Figure 3.11: GO analysis of Embryo-derived TSC-specific CDX2 binding sites generated by accessibility in mESCs.** Bar charts showing -log10 of the hypergeometric p-value GO terms called in highly accessible (ETBS-C1), accessible (ETBS-C2) and inaccessible (ETBS-C3) CDX2 binding sites in mESCs. a,d,g = GO biological processes, b,e,h = MGI mouse phenotype, c,f,i = MGI expression.

Motif analysis suggested that separating ETBS based on accessibility in mESCs might generate distinct peaksets that reflect the differential roles of CDX2 binding. To address this, GO analysis was performed using the single closest gene to all peaks in each cluster using GREAT (McLean et al., 2010). CDX2 binding sites that are 'highly accessible' in mESCs (cluster ETBS-C1) are enriched around genes related to chromatin organization (Figure 3.11a) and those expressed in, and fundamental to, early developmental stages (Figure 3.11b-c). Therefore, unsurprisingly, these 'highly accessible' sites in mESCs likely play a key role during early embryonic development. CDX2 binding here is likely opportunistic.

CDX2 binding sites that are 'accessible' in mESCs (ETBS-C2) are found near genes that important to the trophoblast lineage, including placental development (GO:0001890) and trophectodermal cell differentiation (GO:0001829) (Figure 3.11d). Previous work has shown that *Cdx2* is essential for TSC self-renewal (Strumpf et al., 2005, Niwa et al., 2005). In agreement, these ETBS-C2 sites are near genes associated with stem cell maintenance (GO:0019827) and various abnormal trophoblast phenotypes (Figure 3.11d-e). Similar to ETBS-C1, ETBS-C2 CDX2 sites are concentrated around genes that are expressed at early embryonic stages (Figure 3.11f). Combined with the enrichment of TEAD motifs within these sites, these peaks may represent some of those that CDX2 binds during trophoblast establishment to reinforce commitment to the lineage.

Unlike 'highly accessible' (ETBS-C1) and 'accessible' (ETBS-C2) sites, CDX2 binding sites that are 'inaccessible' in mESCs (ETBS-C3) are not enriched near genes with an obvious trophoblast-related function (Figure 3.11g). Instead, ETBS-C3 are enriched around genes associated with absent trophectoderm (MP:0012102) (Figure 3.11h) and those expressed in later trophoblast lineages (Figure 3.11i). In summary, CDX2 binding sites that are 'inaccessible' in mESCs are enriched at genes related to later trophoblast stages and consensus CDX2 binding motifs are highly prevalent. This agrees with the hypothesis that these sites are unique to, and functional in, the established trophoblast lineage.

## 3.6 Summary

In this chapter, I have characterised three trophoblast stem cell (TSC) lines. Whilst TSC markers such as *Cdx2*, *Eomes* and *Gata3* decrease in expression as cell density and/or time increases, others, such as *Tfap2c* and *Elf5*, do not. This finding is intriguing given that precise ratios of *Eomes*, *Elf5* and *Tfap2c* expression are required to maintain the stem cell fate. When *Elf5* or *Tfap2c* expression increases and/or *Eomes* expression decreases, the balance between these factors are affected and TSC are driven to differentiate (Latos et al., 2015b). As well as informing the density at which cells were collected for CDX2 ChIP-seq, understanding *Cdx2* expression was crucial to design and control *Cdx2* knockdown experiments performed later in this thesis (Chapter 4.1).

Where CDX2 binds in TSCs is different depending on the way the cells were derived. K-means clustering separated CDX2 binding sites into three categories: universal, embryo-derived TSC-specific and transdifferentiated TSC-like cell-specific (mediated by *Oct4* repression) binding sites. Whilst most universal and embryo-derived TSC-specific binding sites are accessible in INX TSCs and an independent wild-type TSC line, TSC-like cell-specific peaks are not. Further, despite being generated by repressing *Oct4* expression, TSC-like cell-specific peaks are also inaccessible in wild-type mESCs and in mESCs 24-hours after *Oct4* repression. GO analysis of these transdifferentiated TSC-specific peaks showed no enrichment of any relevant trophoblast specific term, despite the high proportion of peaks containing the CDX2 consensus motif. Therefore, whether CDX2 binding sites in transdifferentiated TSC-like cells are biologically relevant or functional is unknown.

CDX2 binding sites that were universal or specific to embryo-derived TSCs were located around genes enriched in absent or abnormal trophoblast GO terms. These sites were differentially accessible in mESCs and a subset of these sites increase in accessibility in mESCs after *Oct4* repression. I therefore theorised that it might be possible to use the accessibility of CDX2 binding sites in mESCs to generate peaksets that reflected their functionality. All CDX2 binding sites in embryo-derived TSCs were separated into three new clusters based on their accessibility in mESCs: highly accessible, accessible and inaccessible.

mESCs accessible CDX2 binding sites were enriched around genes expressed in the early embryo and trophoblast lineages and were enriched for TEAD, homeobox, ESRRB and SOX motifs. *Tead4* is upstream of *Cdx2* during trophectoderm establishment and may act as a co-factor for CDX2 during establishment (Nishioka et al., 2009, Rayon et al., 2014, Latos et al., 2015a). Therefore, I hypothesise that these mESC-accessible sites represent initial CDX2 binding targets in the trophoblast lineage that persist in established trophoblast stem cells. Additionally, the ESRRB motif enrichment at these sites agrees with previous work showing that the *Cdx2* binding motif is enriched at ESRRB binding sites (Latos et al., 2015a).

CDX2 binding sites that are inaccessible in mESCs are enriched around genes expressed in later trophoblast stages. These CDX2 binding sites are enriched for other transcription factors motifs, including the extraembryonic ectoderm marker ELF5, suggesting that these are later CDX2 binding targets (Latos et al., 2015b).

Therefore, by integrating chromatin accessibility data with CDX2 ChIP-seq data, I have identified different CDX2 binding site subsets. These subsets are hypothesised to either represent those CDX2 binding sites involved in reinforcing commitment to the trophoblast lineage or distinct CDX2 binding targets within established TSCs. As such, my data suggest that *Cdx2* has distinct targets and roles pre- and post-trophoblast establishment.

# Chapter 4. Profiling the effects of short-term Cdx2 knockdown on the transcriptome and chromatin landscape of Trophoblast Stem Cells

Much work has been carried out over the last 15 years to understand the control of *Cdx2* expression and its role in trophectoderm commitment. This has predominantly used the transdifferentiation of mESCs into TSC-like cells by *Cdx2* overexpression as a model system (Niwa et al., 2005, Kuckenberg et al., 2010, Rhee et al., 2017), while other studies have aimed to understand *Cdx2* in embryo-outgrowth derived TSCs. So far, it has been shown that *Cdx2* is essential for maintaining undifferentiated TSC-like cells (Niwa et al., 2005), perhaps by regulating key TSC marker genes including *Elf5*, *Eomes* and *Esrrb* (Latos et al., 2015a). Much of the rest of the work to understand *Cdx2* in the established trophoblast has centred on generating CDX2 ChIP-seq libraries (Adachi et al., 2013, Chuong et al., 2013). Unfortunately, as I show in Chapter 3, re-analysis of these data suggests that these published CDX2 ChIP-seq libraries are poor in quality and complexity. I also showed that CDX2 binding sites differ in TSC-like cells and embryo outgrowth-derived TSCs. These differences demonstrate why it is important to understand transcription factor binding in the context of the transcriptome and the chromatin landscape.

CDX2 binding sites in TSC-like cells prove not to be accessible in TSCs or mESCs, despite being transdifferentiated from mESCs. In contrast, some, but not all, CDX2 binding sites in embryo-derived TSCs are accessible in mESCs. When analysed based on this differential accessibility in mESCs, it was discovered that ‘accessible’ CDX2 binding sites in mESCs are located around genes expressed at earlier stages of embryonic development than ‘inaccessible’ binding sites. ‘Accessible’ and ‘inaccessible’ CDX2 binding sites in mESCs are also enriched for different motifs, suggesting that at some sites CDX2 binding is opportunistic or occurs through recruitment by other factors, but at others CDX2 recognises its own consensus motif. With this in mind, I hypothesised that different CDX2 peaksets may reflect differential roles of *Cdx2* during trophectoderm fate reinforcement and in established TSCs. Although the differential enrichment of both consensus motifs and GO terms at these sites is highly suggestive, the hypothesis that *Cdx2* plays different roles at different

trophoblast stages assumes that CDX2 binding near a gene is relevant to that gene. However, its binding near a gene, especially in distal regions, is neither indicative of its relevance nor function and requires functional validation (Inoue et al., 2017).

In this chapter I explore the functionality of CDX2 binding in established embryo-derived TSCs. To this end, I generated RNA-seq and ATAC-seq libraries in cells expressing wild-type and reduced levels of *Cdx2* to determine the effect of *Cdx2* loss on the transcriptome and chromatin landscape. By integrating chromatin landscape, transcriptome and CDX2 binding sites, I aimed to define the functional consequences of CDX2 binding. Multiple published histone ChIP-seq datasets were integrated into my analyses to complement libraries generated in-house.

## 4.1 Optimisation of *Cdx2* knockdown

To minimise the effect of changes in *Cdx2* expression during cell culture (Chapter 3.2), all experiments were controlled for starting cell number and culture time.

Previous work has shown that *Cdx2*-null embryos do not form trophoblast outgrowths (Strumpf et al., 2005) and *Cdx2* expression is required to maintain proliferating TSC-like cells (Niwa et al., 2005). As discussed in Chapter 3.4, CDX2 binding sites are enriched around genes associated with abnormal and absent trophoblast lineage GO terms. With these observations in mind, I concluded that the best approach to modulate *Cdx2* transcript levels was to use short-hairpin RNAs (shRNAs). shRNAs are synthetic RNA molecules that are designed to mimic the hairpin precursors of short temporal RNAs used by the intrinsic RNA interference machinery (Paddison et al., 2002). For this, I tested the efficiency of five stem-loop-stem shRNA vectors (generated by The RNAi Consortium (TRC)) that target *Cdx2* (Chapter 2, Table 2.1).

Despite optimisation, the maximal transfection efficiency I achieved in TSCs was approximately 30%. To select for transfected cells, I cloned an mCherry-IRES-puromycin cassette into these shRNA plasmids (Chapter 2.2.7.2). mCherry positive (transfected) and mCherry negative (untransfected) cells were sorted from the population by fluorescent activated cell sorting (FACS) (Figure 4.1a). As a control, a scrambled shRNA vector (SCR) was used in parallel with all shRNA knockdown experiments.

When transfected with any single shRNA vector, *Cdx2* knockdown efficiency was less than 50% (data not shown), but this was improved by co-transfection. When transfected together, shRNA1, shRNA4 and shRNA5 reproducibly knocked down *Cdx2* expression at the transcript and protein level by 65% and 55%, respectively (Figure 4.1b-c).

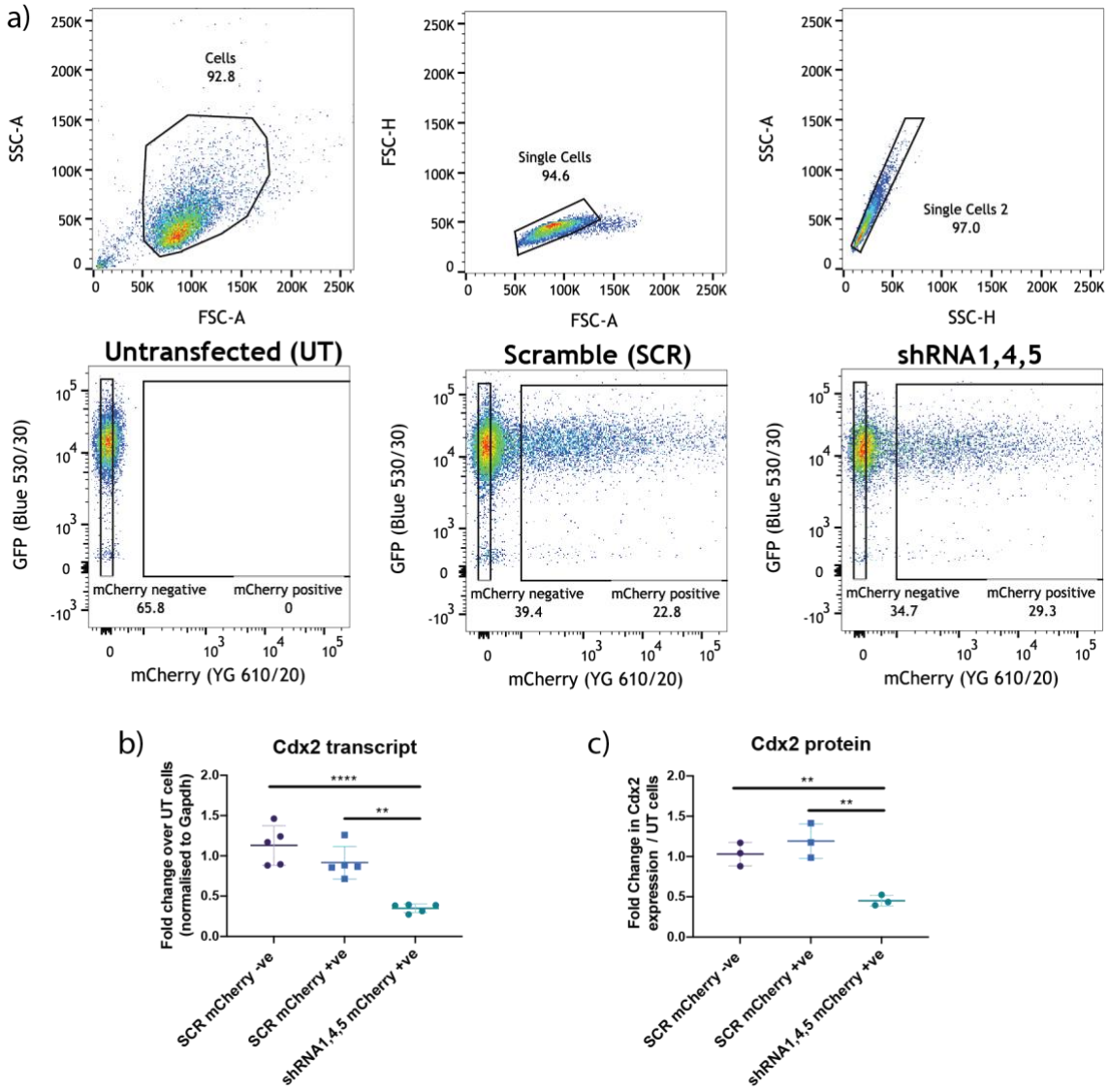



**Figure 4.1: Optimisation of *Cdx2* knockdown in TSCs.** a) Dot plots showing the gating strategy to isolate cell populations based on mCherry expression by FACS. Three sequential gates were applied to stringently isolate single cells: 'cells', 'single cells' and 'single cells 2'. mCherry expression was visualised against GFP (autofluorescence) and gates were drawn to isolate mCherry positive and mCherry negative populations. b) *Cdx2* transcript levels in SCR mCherry negative (-ve), SCR mCherry positive (+ve) and shRNA 1,4,5 mCherry positive (+ve) were analysed by qRT-PCR. Graphs shows fold change in *Cdx2* expression (normalised to *Gapdh*) over untransfected (UT) cells. Significance was determined by one-way ANOVA statistical test: ** p<0.0021, **** p=<0.0001. Error bars show SD, n=5 biological replicate in INX TSCs. c) CDX2 protein level in SCR mCherry negative (-ve), SCR mCherry positive (+ve) and shRNA 1,4,5 mCherry positive (+ve) INX TSCs was analysed by flow cytometry. Graph shows fold change in geometric mean CDX2 expression over geometric mean in UT cells. Significance was determined by one-way ANOVA statistical test: ** p=<0.0021. Error bars show SD (n=3 biological replicates in INX TSCs).

## 4.2 Approach, validation and library preparation

To examine how the loss of *Cdx2* affects the transcriptome and chromatin landscape of established TSCs, RNA-seq and ATAC-seq libraries were generated in wild-type and *Cdx2* knockdown TSCs 48-hour post-transfection. To account for biological variability between samples and to allow for downstream statistical analyses, three biological replicate ATAC-seq and RNA-seq libraries were generated in the following conditions: untransfected (UT), mCherry positive *Cdx2* shRNA knockdown (KD) and scramble mCherry negative (SCR NEG) and positive (SCR POS) TSCs (Figure 4.2a).

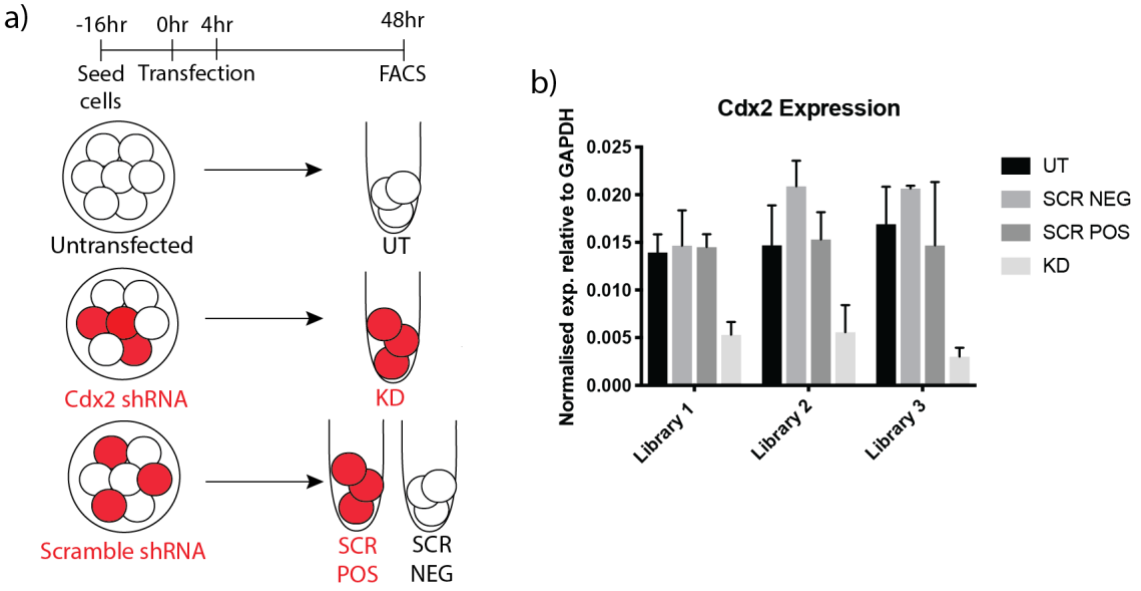



**Figure 4.2: Summary of approach to generate HTS libraries in TSCs expressing wild-type and reduced levels of *Cdx2*.** a) Schematic to illustrate the origins of each condition used to generate RNA-seq and ATAC-seq libraries. Cells were transfected with shRNA constructs for 4 hours approximately 16-hours post-plating. Cells were isolated by FACS based on their mCherry expression as outlined. UT = untransfected, KD = *Cdx2* shRNA knockdown, SCR = scramble. b) qRT-PCR for *Cdx2* expression in RNA samples for each library collected pre-library submission to confirm efficient *Cdx2* knockdown. Error bars display SD for three technical replicates.

Bearing in mind the substantial cost of generating these libraries, I decided to use one TSC line, the INX TSC line, as a core line for generating and validating ATAC-seq and RNA-seq libraries. In Chapter 3, I showed that the INX, XYO and Rossant-GFP TSC lines are similar in their expression of stem cell markers and in CDX2 binding. The INX line was chosen because the XYO cell line had an increased propensity to spontaneously differentiate during culture and the Rossant-GFP TSCs did not have a passage number.

A total of 50,000 cells was collected using FACS (outlined in Figure 4.2a) for each ATAC-seq and RNA-seq library. Biological replicate libraries were generated in parallel over three collections. qRT-PCR was performed for each RNA-seq sample to confirm efficient *Cdx2* knockdown (Figure 4.2b). *Cdx2* knockdown efficiency was over 65% for all libraries relative to control libraries.

## 4.3 *Cdx2* knockdown in established trophoblast stem cells perturbs the transcriptome

As RNA-seq libraries were generated using a poly(A) enrichment kit, they only contain mature mRNA, rather than total RNA. Read quality and alignment efficiency for all libraries was high, with >98% of all reads mapped in each library (data not shown). An initial principal component analysis (PCA) was performed to assess similarities and differences between all libraries generated in an unbiased manner (Figure 4.3a). This revealed that most variance observed between libraries, as shown in PC1 (54%), is due to biological variation between replicates. Despite this biological heterogeneity, PC2 (24%) clearly separates the KD samples from all of the control samples. Therefore, *Cdx2* knockdown perturbs the transcriptome of TSCs.

Pairwise differential gene expression analysis using DESeq2 revealed that 1138, 1370 and 964 genes were dysregulated ($p<0.05$) in KD versus UT, SCR NEG and SCR POS, respectively. The log2 fold change for the majority of these genes was low (all less than 2.8-fold) as shown in the MA plots in Figure 4.3b-c. Small fold changes in expression levels are unsurprising, bearing in mind that the overall knockdown efficacy is only 53-62% (Figure 4.3d). These low fold-changes make it more challenging to create a stringent list of genes dysregulated in *Cdx2* knockdown.

To make a list of dysregulated genes, I decided to overlap those genes that were differentially expressed (adjusted $p<0.05$) in KD versus control pairwise comparisons. Although SCR NEG samples cluster well with SCR POS and UT samples (Figure 4.3a), *Cdx2* knockdown efficiency in KD versus SCR NEG (53%) was poorer than anticipated (Figure 4.3d). Therefore, genes that were differentially expressed in the same direction in KD versus SCR POS (herein after referred to as SCR) and UT were used to generate representative gene lists without SCR NEG. No fold change criteria were implemented to generate this final gene lists (Figure 4.3e). In total, 263 genes were downregulated and 363 genes were upregulated by *Cdx2* knockdown.

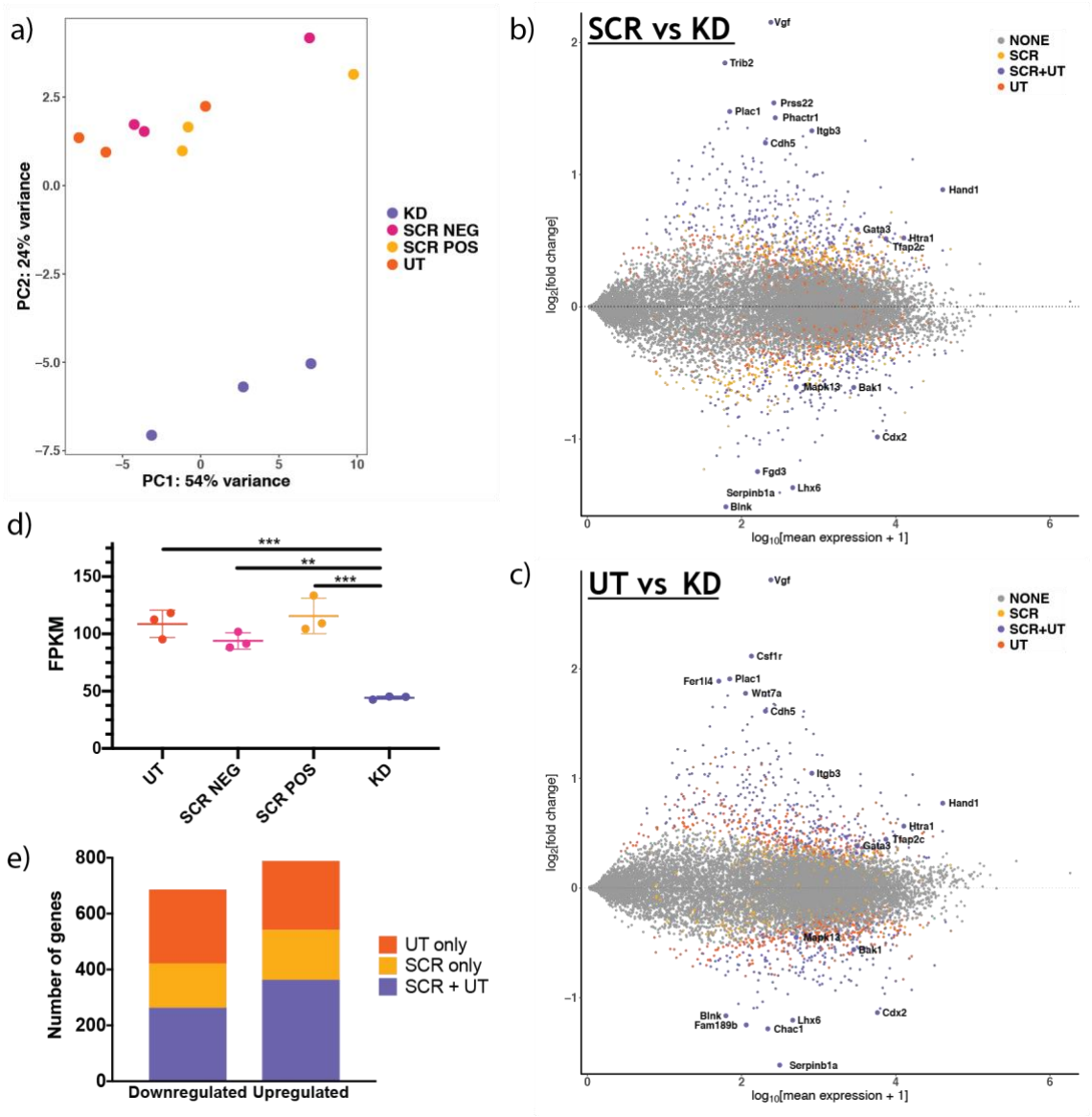



**Figure 4.3: 48-hour *Cdx2* knockdown significantly perturbs the TSC transcriptome.** a) Principal component analysis (PCA) of RNA-seq data generated using ggplot (version 2.2.1) within the R programming environment (version 3.3.1). Three biological replicates were generated for *Cdx2* shRNA knockdown (KD), scramble transfected (SCR POS), scramble untransfected (SCR NEG) and untransfected (UT) TSC libraries. b-c) MA plot generated using ggplot (version 2.2.1) within the R programming environment (version 3.3.1) to show transcript fold change against normalised mean expression for all genes. In (b), the $\log_2$ fold changes of SCR (SCR POS) versus KD are shown and (c) shows the $\log_2$ fold changes of UT versus KD. Genes in grey are not differentially expressed. Genes in yellow and orange are only differentially expressed in SCR or UT versus KD, respectively. Purple genes are those that are differentially expressed in both pairwise comparisons. d) Fragments per kilobase of exon model per million reads mapped (FPKM) for *Cdx2* in each library (n=3 per condition). Significance determined by one-way ANOVA: ** p=0.0019, *** p≤0.0002. e) Bar chart summarising differentially expressed genes in both SCR or UT versus KD pairwise comparison as outline in the MA plots in (c-d).

Next, I compared the expression of dysregulated genes in all RNA-seq libraries. Although they were not used to make this gene list, SCR NEG replicates express high levels of downregulated genes and vice versa, validating both this gene list and

the approach taken (Figure 4.4a). This heatmap also revealed that SCR R2 was an outlier, expressing lower levels of most genes downregulated in *Cdx2* knockdown. Having defined what genes are affected by *Cdx2* knockdown, I aimed to understand its effect on the whole transcriptome. As the lowest expressed dysregulated genes in the final gene list were 0.0825 transcripts per million (TPM) across all samples (for *Cd93* and *Nwd1*), genes with an average TPM of ≥0.08 were deemed to be expressed (13418). By this definition, 4.7% of all genes expressed in TSCs are significantly perturbed as a result of *Cdx2* knockdown (Figure 4.4b).

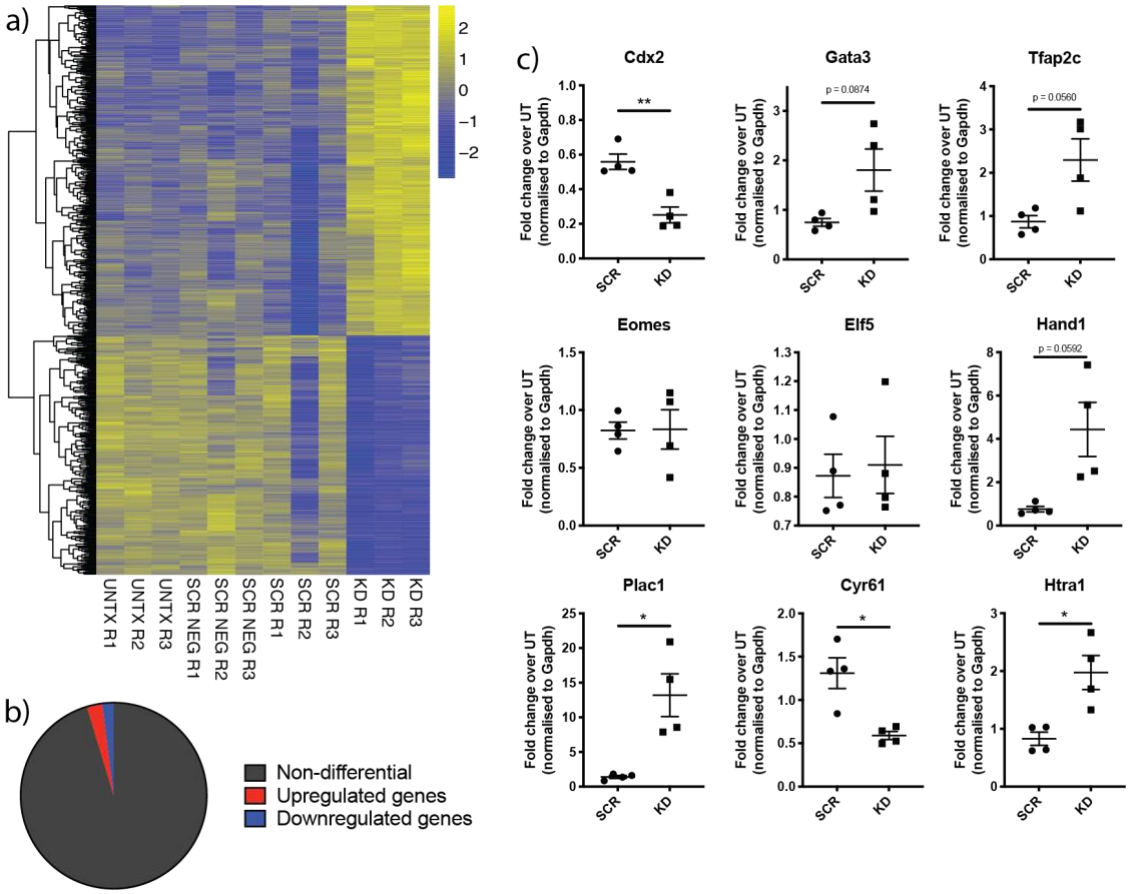



**Figure 4.4: Validation of genes differentially expressed by Cdx2 knockdown.** a) Heatmap displaying $\log_2$ fold change in expression of differentially expressed genes over the average expression of each gene across all samples. Plot generated using ggplot (version 2.2.1) within the R programming environment (version 3.3.1). b) Pie chart to show the proportion of expressed genes, defined as being expressed on average at ≥0.08 TPM, that are up- and downregulated in *Cdx2* knockdown. c) qRT-PCR analysis of differentially expressed genes known to be expressed in the trophoblast lineage. Graphs displayed as fold change (gene expression normalised to *Gapdh*) over UT samples. Significance determined by t-test: * p≤0.0332, ** p≤0.0021. Error bars show SEM. n=4 biological replicates.

To validate RNA-seq findings, I performed qRT-PCR on a subset of differentially expressed genes. Only two trophoblast stem cell marker genes were affected by

short-term *Cdx2* knockdown: *Gata3* and *Tfap2c* (Figure 4.4c). The upregulation of *Gata3* and *Tfap2c* agrees with previous work suggesting that *Cdx2* expression is downstream of these factors in established TSCs (Home et al., 2009, Kidder and Palmer, 2010). In contrast, *Cdx2* is thought to be upstream of two other TSC marker genes, *Eomes* and *Elf5*, during trophoblast specification (Niwa et al., 2005, Ng et al., 2008) and all three co-activate one another within established TSCs (Chuong et al., 2013, Latos et al., 2015b). As neither *Eomes* nor *Elf5* expression are affected by short-term *Cdx2* knockdown, they are not direct targets of, or not responsive to, *Cdx2* in TSCs (Figure 4.4c). Other TSC markers such as *Tead4*, *Ets2* and *Sox2* were not differentially expressed as these were not included in the final gene list. Therefore, their expression was not assessed by qRT-PCR. Although few TSC marker genes are affected by *Cdx2* knockdown, other trophoblast associated genes such as *Plac1* (Jackman et al., 2012), *Htra1* (Nie et al., 2005), *Cyr61* (Mo et al., 2002) and the trophoblast giant cell marker gene *Hand1* (Riley et al., 1998) are perturbed by *Cdx2* knockdown (Figure 4.4c).

To understand the potential functions of the genes dysregulated by *Cdx2* knockdown, gene ontology (GO) analysis was performed using MetaCore (Figure 4.5). In contrast to *Cdx2*-null trophectoderm, which shows increased levels of apoptosis (Strumpf et al., 2005), short-term *Cdx2* knockdown caused the downregulation of genes associated with apoptosis, inflammation and the immune response (Figure 4.5a). Reduced expression of two genes found in several of these GO terms, *Mapk13* and *Bak1*, was validated by qRT-PCR (Figure 4.5b). Genes that were upregulated by *Cdx2* knockdown are associated with GO terms related primarily to cell adhesion and other developmental terms (Figure 4.5c). Two genes that are found in several of these upregulated gene GO terms, *Cdh5* and *Itgb3*, were validated as upregulated in KD cells (Figure 4.5d).

I conclude that *Cdx2* knockdown perturbs the transcriptome, but the biological relevance of the affected genes, whether upregulated or downregulated, is unclear.

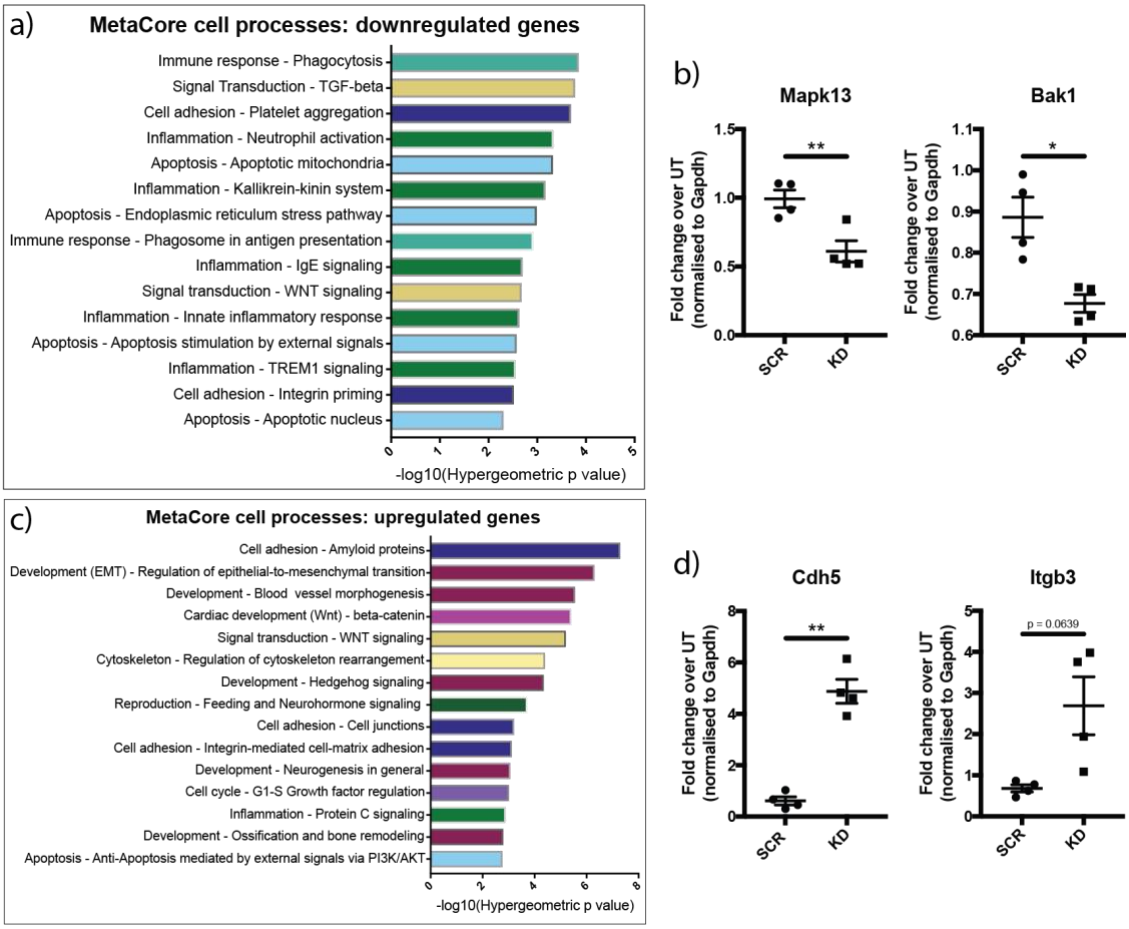



**Figure 4.5: MetaCore cell processes enrichment analysis for upregulated and downregulated genes.** a,c) MetaCore cell processes analysis for genes that are downregulated and upregulated genes resulting from *Cdx2* knockdown respectively. b,d) qRT-PCR for key genes included in multiple differentially enriched terms in a and c. Significance was determined using t-tests: * $p \leq 0.0032$, ** $p \leq 0.0021$. Error bars display SEM, n=4 biological replicates.

## 4.4 *Cdx2* knockdown alters chromatin accessibility across the chromatin landscape

To understand how *Cdx2* knockdown affects the chromatin landscape I generated ATAC-seq data in parallel with the RNA-seq libraries discussed above. ATAC-seq uses a hyperactive Tn5 transposase that integrates adapter sequences preferentially in regions of open, accessible chromatin (Buenrostro et al., 2013). These regions can be PCR-amplified and sequenced to generate libraries that display overall chromatin accessibility within a relatively small starting pool of cells.

To assess in an unbiased fashion the way in which UT, SCR, SCR NEG and KD ATAC-seq libraries relate to one another, principal component analysis was performed. There is some variation between libraries (PC2 = 15%), but most differences in the chromatin landscape are caused by *Cdx2* knockdown (PC1 =72%) (Figure 4.6a). This effect of *Cdx2* knockdown on the chromatin landscape must be at specific loci as the correlation between all samples was high, both genome-wide (Figure 4.6b) and at individual loci (Figure 4.6c).

These ATAC-seq libraries have the characteristic visual profiles of ATAC-seq libraries, with peaks denoting regions of accessible chromatin around the *Cdx2* locus (Figure 4.6c). Biological replicates for each sample were merged to generate overall alignment files for each condition (UT, SCR NEG, SCR and KD). As expected, all of these ATAC-seq libraries have a normal fragment length distribution, with the majority of sequenced reads either nucleosome-free or of mononucleosome-associated lengths (Buenrostro et al., 2013) (Figure 4.6d). I conclude that all libraries are of good quality and can be used to assess the functional effect of *Cdx2* knockdown at the chromatin level.

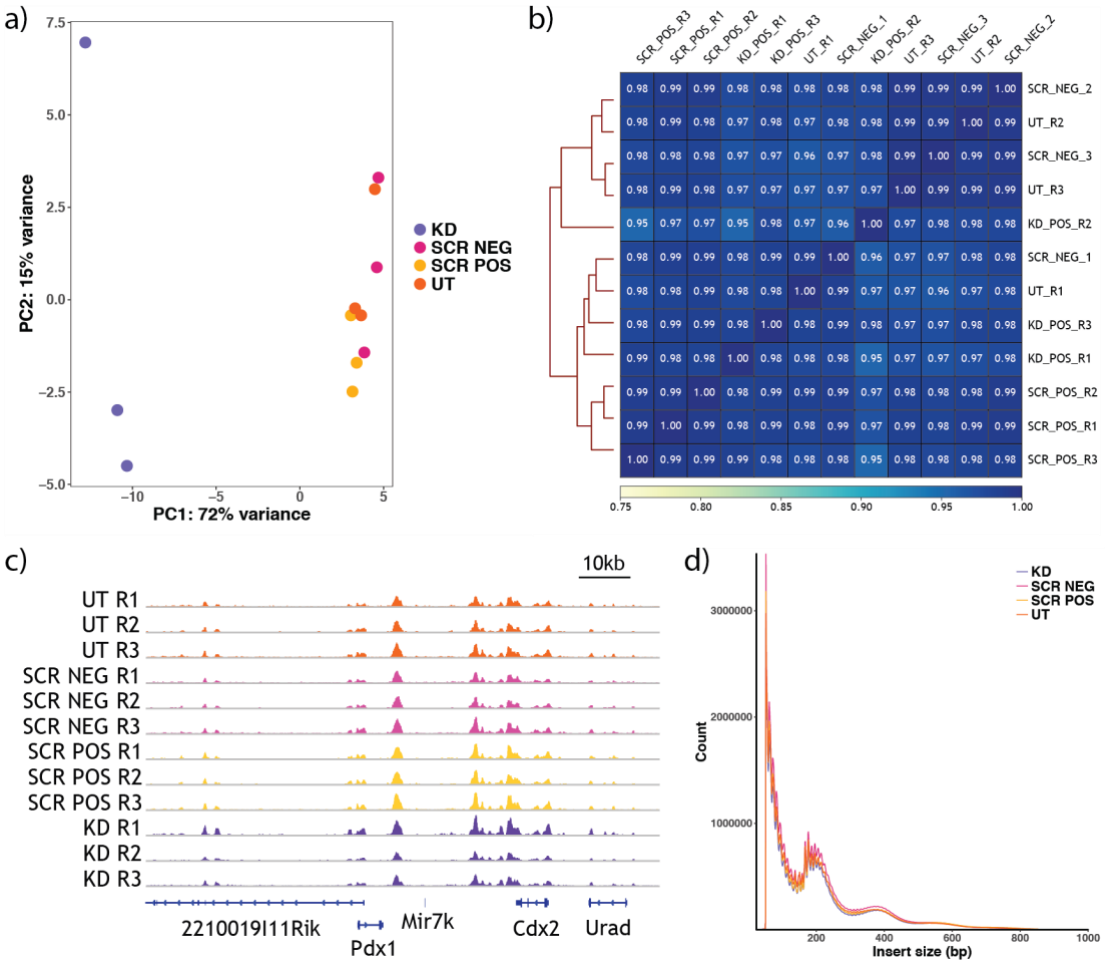



**Figure 4.6: Quality control assessment of ATAC-seq libraries.** a) Principal component analysis (PCA) generated using ggplot (version 2.2.1) within the R programming environment (version 3.3.1) for ATAC-seq libraries generated in biological triplicate. Those conditions assessed are *Cdx2* shRNA knockdown (KD), scramble transfected (SCR POS), scramble untransfected (SCR NEG) and untransfected (UT) TSCs. Read quality and alignment efficiency was high, with less than 30% of all reads displaying mitochondrial contaminants and low duplication rates (data not shown). b) A correlation heatmap showing Pearson correlation coefficients. This was calculated using read density at 10,000bp intervals across the genome. c) IGV track displaying accessibility for each biological replicate sample individually around the *Cdx2* locus. d) Insert size plot displaying the lengths of paired-end alignments in merged samples per condition.

Regions of differential accessibility were determined by pairwise analysis for KD against SCR and UT. SCR NEG samples were discounted from this analysis because of lower *Cdx2* expression in the corresponding RNA-seq libraries. To generate a consensus peakset of chromatin regions affected by *Cdx2* knockdown, only peaks with p<0.01, fold change 0.25x and those present in both pairwise comparisons were included in the final peakset. Regions were discounted if they were found to differ between control (SCR and UT) samples.

In total, there were 2313 (37.35%) and 3879 (62.65%) sites of lost accessibility (LA) and gained accessibility (GA), respectively (Figure 4.7a). Differences in accessibility

at these sites between wild-type TSCs and *Cdx2* knockdown TSCs were validated globally using heatmaps (Figure 4.7b) and locally around genes dysregulated by *Cdx2* knockdown, including *Cdh5*, *Plac1* and *Cyr61* (Figure 4.4, Figure 4.7c-e) (Thorvaldsdottir et al., 2013). Overall, these data indicate that a short-term reduction in *Cdx2* expression by only 53-62% alters the chromatin landscape at specific loci.

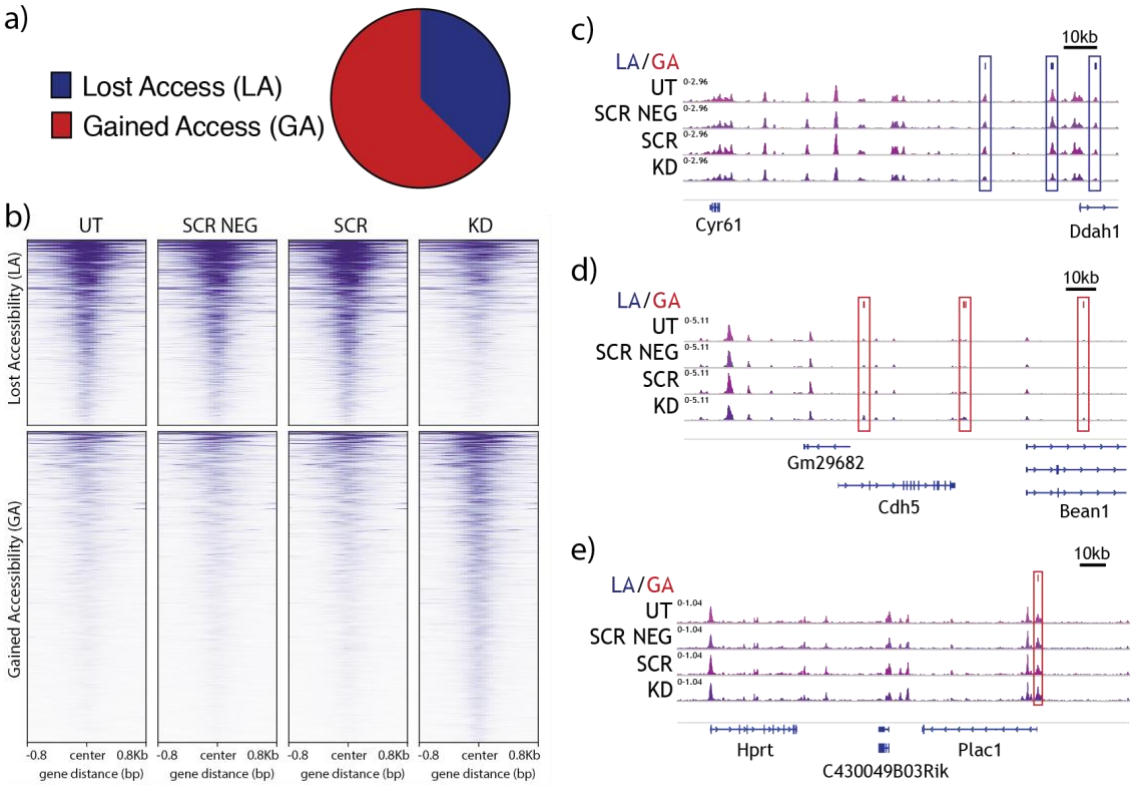



**Figure 4.7: Short-term *Cdx2* knockdown causes significant changes to chromatin accessibility.** a) Pie chart to show the relative proportion of lost accessibility (LA) and gained accessibility (GA) sites resulting from *Cdx2* knockdown. b) Heatmap to show normalised read density of UT, SCR NEG, SCR and KD ATAC-seq libraries at all LA and GA sites. c-e) IGV track to show sites of differential accessibility at the downregulated gene *Cyr61* locus and upregulated genes *Cdh5* and *Plac1* loci. Boxes highlight those regions called as differentially accessible in KD libraries relative to WT libraries.

To discover which transcription factors contribute to chromatin accessibility changes, I next performed motif and footprinting analyses. Transcription factors often bind to conserved DNA binding consensus motifs (Khan et al., 2018, Mathelier et al., 2016). There are several algorithms that can scan defined sets of chromatin intervals for consensus motifs and determine the transcription factors that may bind them (Heinz et al., 2010, Khan et al., 2018, Bailey et al., 2009). Although not the case for all transcription factors, their binding to motif sequences can protect the region from enzymatic digestion (Galas and Schmitz, 1978). In this way, the region can be locally

protected, leading to a ‘footprint’ of decreased read coverage around motifs (Galas and Schmitz, 1978). However, not all transcription factors have a footprint and in, some cases, signatures arise from biases in cutting rather than transcription factor binding (He et al., 2014, Grontved et al., 2015, Sung et al., 2014, Swinstead et al., 2016). To address and minimise the effect of these issues, the tool BaGFoot integrates two different transcription factor-DNA binding phenomena to perform differential footprinting analysis: footprinting and accessibility around motifs between conditions, referred to as flanking accessibility (Baek et al., 2017).

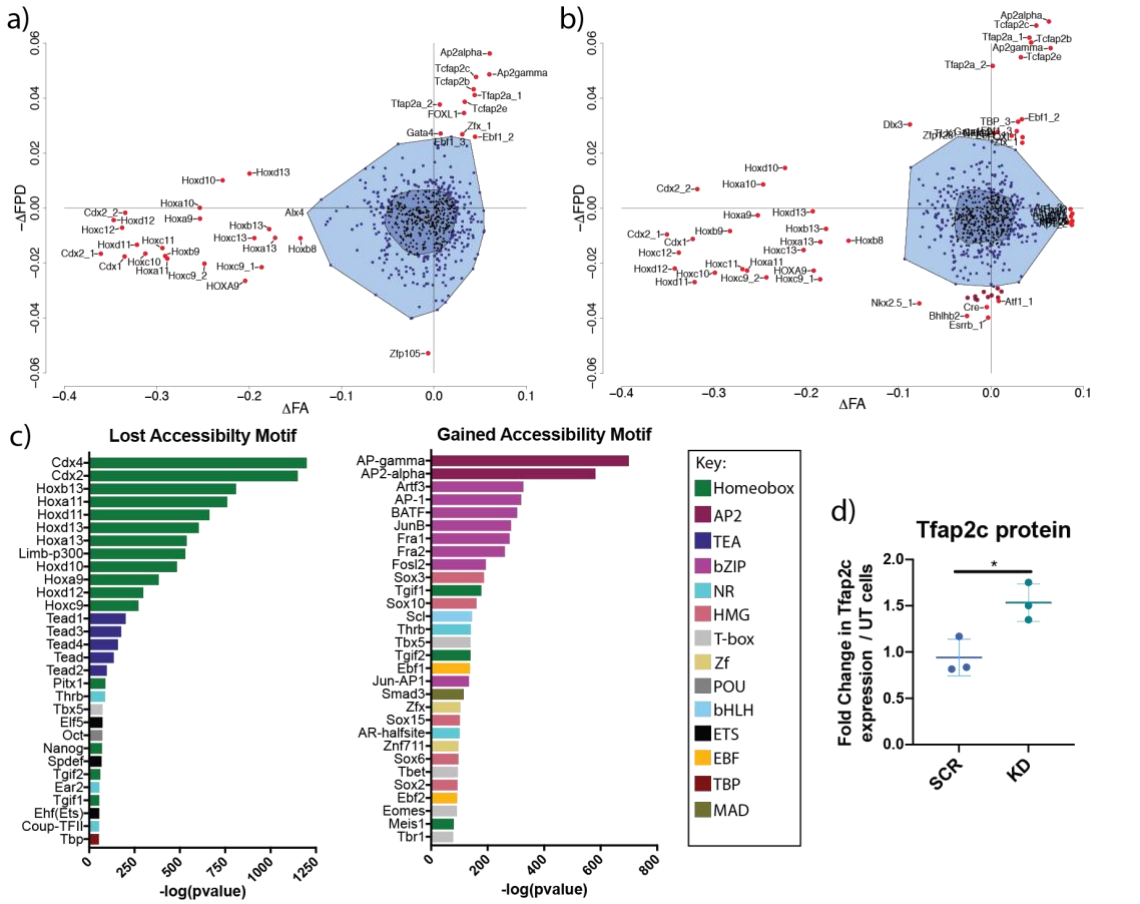



**Figure 4.8: Sites of lost and gained accessibility are attributable to CDX2 and TFAP2C, respectively.** a-b) Bag plot depicting those motifs enriched as having a differential footprint depth (ΔFPD) and flanking accessibility (ΔFA) in KD versus SCR (a) and UT (b). Motifs that are statistically significant outliers are shown as red circles. c) Bar charts displaying the 30 most statistically significant enriched motifs in lost accessibility and gained accessibility peaksets as determined using HOMER. d) Fold change in TFAP2C protein expression in KD and SCR over UT as assessed by flow cytometry. Error bars show SD. n=3 biological replicates. Statistical significance determined by t-test: * = 0.0223.

Footprinting analysis using BaGFoot attributed the loss and gain of accessibility in *Cdx2* knockdown conditions over control conditions (SCR and UT) to the homeodomain CDX/HOX and AP2 motif families, respectively (Figure 4.8a-b) (Baek et al., 2017). The enrichment of these motifs was confirmed by conventional motif analysis using HOMER (Figure 4.8c) (Heinz et al., 2010). Overall, 69% and 76% of LA sites contained the CDX2 and CDX4 consensus motifs, respectively. This suggests the loss of accessibility in *Cdx2* knockdown is a direct consequence of decreased CDX2 expression and its binding at these sites. The TSC marker gene *Tfap2c*, an AP2 transcription factor family member, is upregulated in *Cdx2* knockdown cells at both the transcript and protein levels (Figure 4.4c, Figure 4.8d) and is probably the main cause of gained accessibility in *Cdx2* knockdown.

I next determined where GA and LA sites are distributed across the genome (Heinz et al., 2010). Most GA and LA sites are located in introns or are intergenic (Figure 4.9a-b). Using GREAT, I performed GO analysis using the single closest gene TSS to every LA and GA site (McLean et al., 2010). To make this as biologically relevant as possible, enrichment analysis was performed over a background file containing all accessible regions in all ATAC-seq libraries (Chapter 2.5.2).

Limited GO terms were returned for LA sites (Figure 4.9c), perhaps because these sites are few in number and/or almost half are located over 10kb away from the closest gene. Of those GO terms enriched for LA sites, several were related to cell-cell junctions: apical junction complex (GO:0043296) and tight junction (GO:0005923) (Figure 4.9d). This aligns with the *in vivo* phenotype of *Cdx2*-null embryos which show abnormal junctions by the late blastocyst stage (Strumpf et al., 2005). Only one trophoblast lineage related MGI expression term was discovered in the top 50: E6.5 (TS5) trophectoderm (ID: 99, $-\log_{10}$(Hypergeometric p value) = 3.40) (Figure 4.9e). Instead, genomic intervals are near genes expressed in other *Cdx2*-expressing tissues including the intestine and tail regions. As such, the functional relevance of LA sites for the trophoblast lineage is not immediately apparent.

In contrast, gene ontology analysis for GA genomic regions suggest that sites are enriched around genes associated with trophoblast lineage derivatives from later stages of embryonic development, mainly E14 (TS22) (Figure 4.9f). The two most

enriched GO biological processes terms are trophoblast giant cell differentiation (GO:0060707) and cell differentiation involved in embryonic placenta development (GO:0060706) (Figure 4.9g). This suggests that the biological function of sites that gain accessibility after *Cdx2* knockdown might be to drive the cells to differentiate.

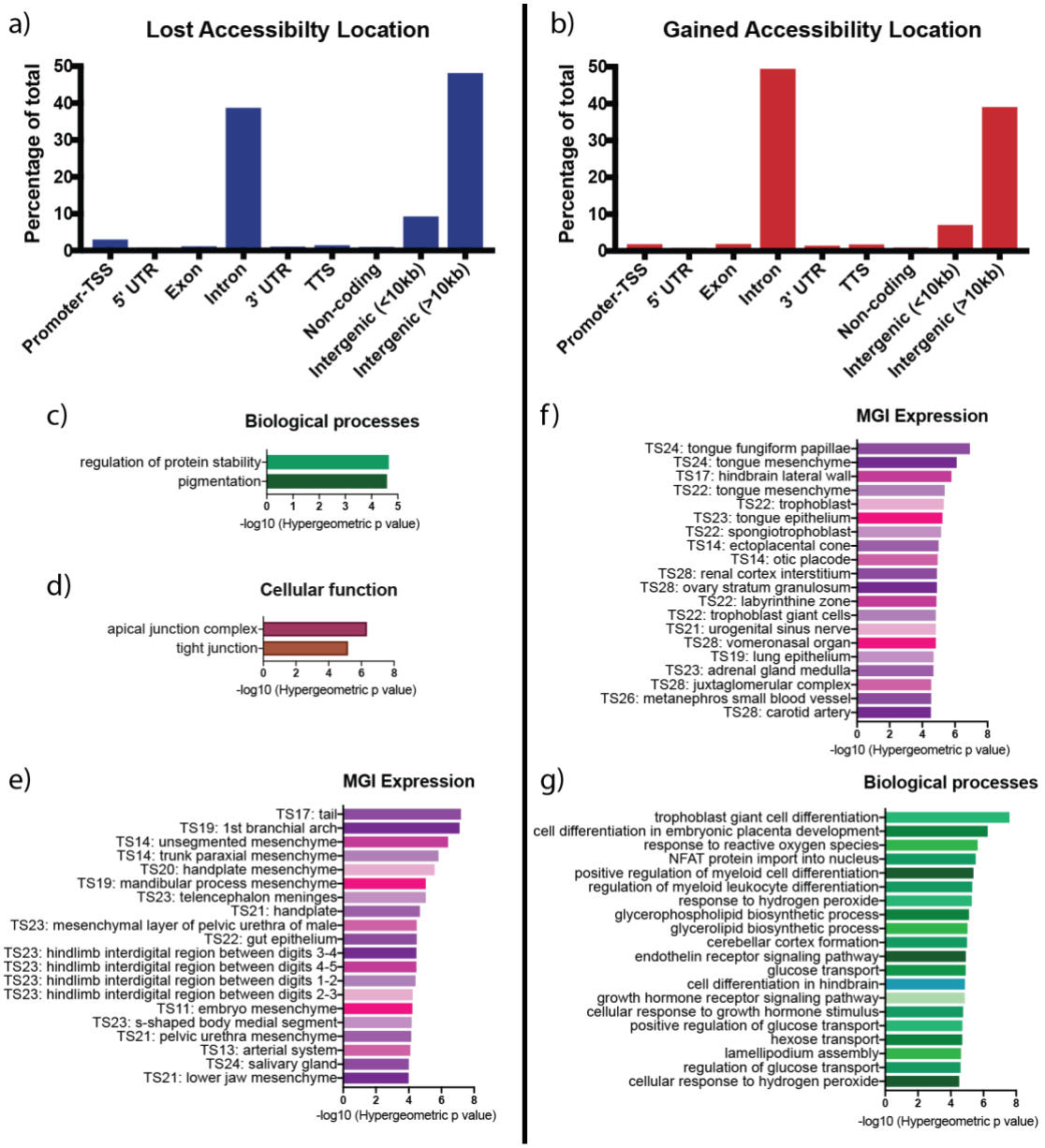



**Figure 4.9: Functional assessment of differentially accessible regions in Cdx2 knockdown by location and GO analysis.** a-b) Location of LA and GA sites displayed as a percentage of the total regions as determined by HOMER annotation. c-g) Bar charts to show -log10 of the hypergeometric p-value GO terms called in LA (left) and GA (right) over total accessible sites in all samples. Analysis was performed using the single nearest gene association rule by GREAT. c,g = GO biological processes, d = GO cellular function, e,f = MGI expression.

## 4.5 Library integration: ATAC-seq, ChIP-seq and RNA-seq

To fully understand the role of *Cdx2* in the established trophoblast lineage, the relationship between CDX2 binding and the effect of *Cdx2* knockdown on the transcriptome and chromatin accessibility need to be integrated.

### 4.5.1 To lose or not to lose: CDX2 binding, knockdown and changes in accessibility

As shown in the previous section, sites of reduced accessibility in *Cdx2* knockdown cells (LA) are enriched for the CDX2 homeodomain motif and footprint. In Chapters 3.3 and 3.4, I discussed the generation and analysis of CDX2 ChIP-seq libraries in embryo-derived TSCs. Most, but not all, lost accessibility sites are bound by CDX2 in these ChIP-seq libraries (Figure 4.10a). To quantify the number of CDX2 binding sites called in ChIP-seq libraries that overlap with sites of reduced accessibility in *Cdx2* knockdown, these peaksets were intersected (Figure 4.10b). Overall, 92% of LA sites directly overlap CDX2 binding sites in one of the CDX2 ChIP-seq libraries. This suggests that only 7.6% of all reduced accessibility sites in *Cdx2* knockdown are unrelated to CDX2 biding.

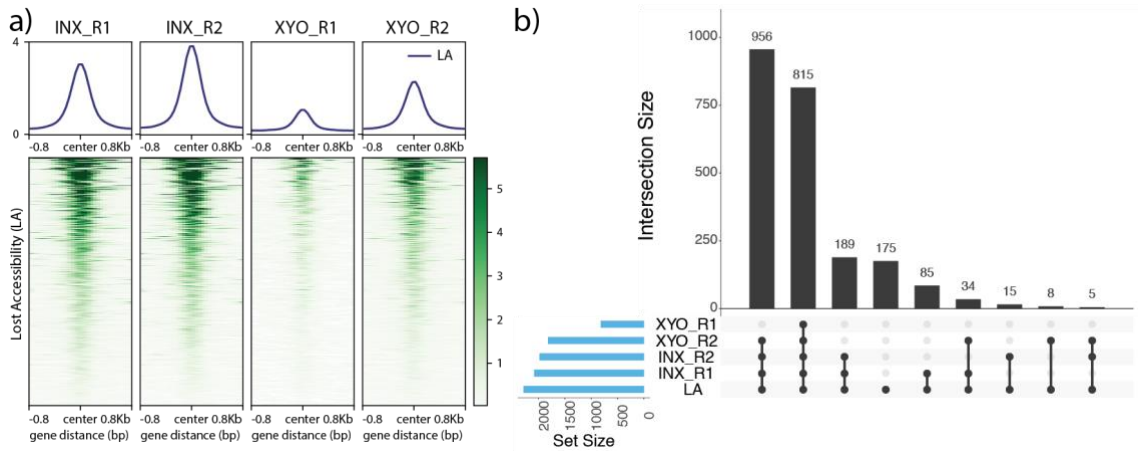



**Figure 4.10: Sites of lost accessibility show significant overlap with called CDX2 binding sites.** a) Normalised read density heatmap to show enrichment of INX and XYO CDX2 ChIP-seq dataset at sites of lost accessibility (LA) by ATAC-seq in *Cdx2* knockdown. b) UpSetR plot showing intersection of genomic intervals contained in LA, INX and XYO datasets.

CDX2 binding sites in TSCs differ according to the way in which the cells were originally derived. In the previous chapter, I showed that CDX2 binding sites in embryo-derived TSCs (ETBS) can be separated based on their accessibility in

mESCs (ETBS-C1, -C2 and -C3) (discussed in Chapter 3.5). Motif and GO analysis suggested that these different ETBS clusters may have different roles in the trophoblast lineage (Chapter 3.5). To discover whether one of these CDX2 binding site clusters is more affected by *Cdx2* knockdown than the others, I generated profile plots that display how accessible they are, on average, in UT, SCR and KD conditions. This revealed that chromatin accessibility is reduced in all these clusters in KD samples relative to the wild-type controls, but there are also differences between SCR and UT (Figure 4.11a). This suggests that all ETBS peaksets may contain CDX2 binding sites that lose accessibility in *Cdx2* knockdown. To determine if one ETBS cluster is more affected than the others, ETBS sites need to be directly overlapped with LA sites.

To address this, I intersected LA peaks with those in the ETBS clusters (Figure 4.11b). Of the sites that lose accessibility in *Cdx2* knockdown cells, 78.4% are contained in cluster ETBS-C3 (Figure 4.11b). The high overlap between ETBS-C3 and LA sites is unsurprising as this ETBS cluster showed the highest enrichment for homeobox motifs (Chapter 3, Figure 3.10c). However, despite this high enrichment for homeobox motifs, only 6.35% of all ETBS-C3 sites are affected by *Cdx2* knockdown.

To understand why accessibility of CDX2 binding sites are only lost at specific sites, I removed CDX2 binding sites that lose accessibility in *Cdx2* knockdown (LA) from each ETBS peaksets (ETBS-C1/2/3-minus LA (-mLA)). The read density heatmap in Figure 4.11c confirmed that the only CDX2 binding site peakset that significantly reduced chromatin accessibility between KD and UT/SCR conditions was LA sites (purple, Figure 4.11c). These LA sites are also those that are most enriched for CDX2 binding in TSCs (green, Figure 4.11c), suggesting that the loss of CDX2 binding at these sites may be because this is where CDX2 binds the strongest. There are, however, many other sites where CDX2 binds robustly whose accessibility is unaffected by *Cdx2* knockdown.

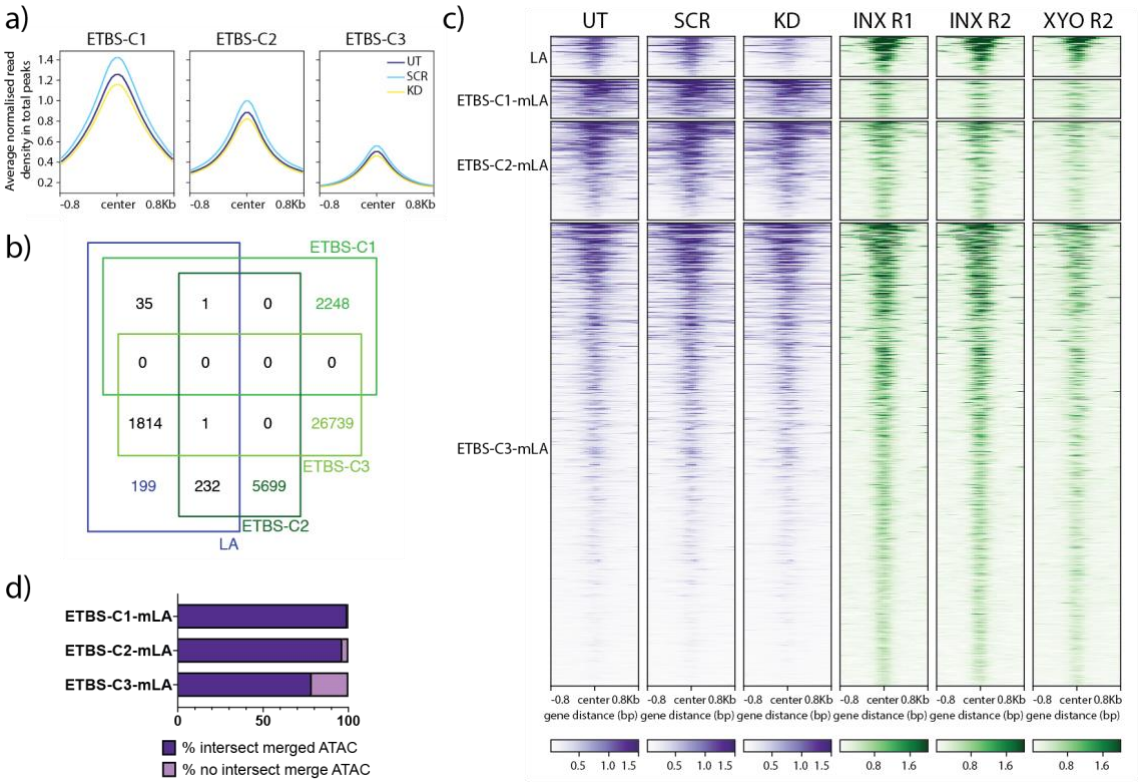



**Figure 4.11: Integration of ATAC-seq and CDX2 ChIP-seq libraries.** a) Profile plot to show the average read density of UT, SCR and KD ATAC-seq libraries at embryo-derived TSC CDX2 binding site (ETBS) clusters generated in Chapter 3.5. b) Venn diagram displaying the number of unique and overlapping peaks between lost accessibility sites (LA) and ETBS clusters. c) Read density heatmap for UT, SCR and KD ATAC-seq libraries (purple) and ChIP-seq libraries (green) generated in INX and XYO cell lines at LA sites and ETBS clusters minus LA sites (-mLA). d) Bar chart showing the percentage of each ETBS cluster minus sites of reduced accessibility in *Cdx2* knockdown (-mLA) that overlaps with a merged file of accessible chromatin across all ATAC-seq libraries, "merged all INX".

Next, I explored why chromatin accessibility did not change at most CDX2 binding sites by examining each ETBS cluster independently. In Chapter 3.5, it was hypothesised that Cluster ETBS-C1 represented opportunistic CDX2 binding sites due to the lack of homeobox motif enrichment and their high accessibility in mESCs (Chapter 3.5, Figure 3.10c). In agreement with this, only 1.6% of ETBS-C1 sites overlap with LA peaks (Figure 4.11b) and CDX2 binding (green, Figure 4.11c) is lower in ETBS-C1-mLA than in other clusters even though these sites are highly accessible in TSCs (purple, Figure 4.11c). The unchanging accessibility at ETBS-C1-mLA sites after *Cdx2* knockdown is, therefore, likely because CDX2 binding here is opportunistic.

As with ETBS-C1, only 3.9% of ETBS-C2 sites overlap with LA sites (Figure 4.11b) and CDX2 binding (green, Figure 4.11c) is low relative to how accessible (purple,

Figure 4.11c) ETBS-C2-mLA sites are. Chapter 3.5, I showed that ETBS-C2 sites are enriched for the TEAD family of motifs, suggesting that CDX2 may be recruited to these sites by TEAD4. The accessibility of ETBS-C2-mLA sites is, therefore, unaffected by *Cdx2* knockdown because CDX2 binding here is likely recruited by other factors and is not required to maintain accessibility.

Although 79.5% of LA sites overlap with CDX2 binding sites in cluster ETBS-C3, 93.7% of ETBS-C3 sites do not show a significant reduction in accessibility after *Cdx2* knockdown (Figure 4.11b). The ETBS-C3 sites are highly enriched for homeobox motifs (Chapter 3.5, Figure 3.10c) as are ETBS-C3 sites after LA sites are removed: ETBS-C3-mLA (data not shown). In contrast to other ETBS clusters, the strength of CDX2 binding (green, Figure 4.11c) in ETBS-C3-mLA sites correlated with how accessible the chromatin was (purple, Figure 4.11c). However, many ETBS-C3-mLA sites show negligible accessibility (Figure 4.11c) and 21.5% of ETBS-C3-mLA sites were not called as accessible by MACS2 in any TSC ATAC-seq library (Figure 4.11d). In contrast, 0.6% and 3.5% of sites were inaccessible in TSCs for ETBS-C1-mLA and ETBS-C2-mLA, respectively (Figure 4.11d).

Given the above, it is likely that the resistance of ETBS-C3-mLA sites to chromatin accessibility changes in *Cdx2* knockdown may occur for different reasons. To address this, ETBS-C3-mLA sites were sub-divided into new clusters based on their accessibility in TSCs. Those 21.5% of ETBS-C3-mLA sites that were not deemed to be accessible by MACS2 peak calling on ATAC-seq libraries were considered to be a new, inaccessible subset. K-means clustering was subsequently performed with the remaining peaks to separate them into new clusters. In all, this generated three new sub-peaksets for ETBS-C3-mLA: highly accessible, accessible and inaccessible (Figure 4.12a). Next, I determined whether these ETBS-C3-mLA clusters were differentially enriched for motifs that may explain why they are unaffected by *Cdx2* knockdown.

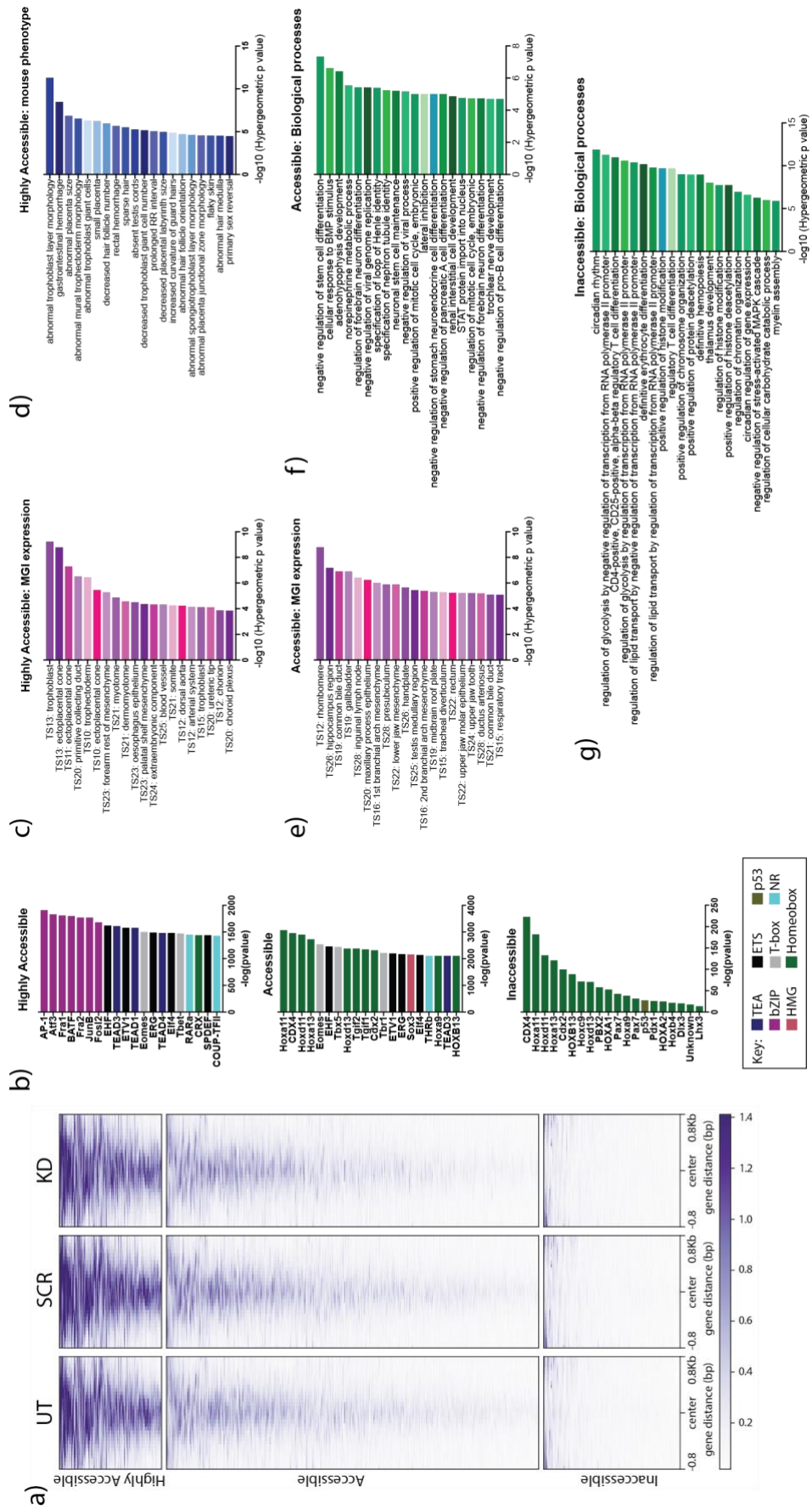



**Figure 4.12: Subsetting of mESC-inaccessible CDX2 binding sites by accessibility in TSCs.** a) A heatmap displaying UT, SCR and KD ATAC-seq read density at CDX2 binding sites inaccessible in mESCs and after the removal of those that sites lose accessibility in *Cdx2* knockdown: ETBS-C3-mLA. Those CDX2 binding sites that were inaccessible in all ATAC-seq libraries were removed prior to k-means clustering. K-means clustering separated the remaining sites into highly accessible and accessible clusters. b) The -log(p value) of enriched motifs called in each new cluster outlined in (a) using HOMER. c-g) Bar charts showing -log10 of the hypergeometric p-value GO terms called in highly accessible, accessible and inaccessible ETBS-C3-mLA clusters using GREAT. c,e = MGI expression, d = MGI mouse phenotype, f,g = GO biological processes

'Highly accessible' ETBS-C3-mLA sites are most enriched for bZIP motifs (Figure 4.12b) and several of these motifs, including JunB and AP-1, are also found in sites with increased accessibility as a result of *Cdx2* knockdown (Figure 4.8c). Whilst homeobox motifs are poorly enriched in 'highly accessible' sites, motifs for other key TSC markers including *Elf5*, *Ets2* (both ETS), *Eomes* and *Tead4* (TEA(D)) are significantly enriched (Figure 4.12b). As both EOMES and TEAD4 are known to directly interact with CDX2 (Latos et al., 2015a), its binding at these sites is probably facilitated by other factors and/or is opportunistic. Unsurprisingly, bearing in mind their high accessibility in TSCs and the enrichment of TSC marker gene motifs, ETBS-C3-mLA 'highly accessible' sites are enriched around genes that are expressed in (Figure 4.12c) and critical to (Figure 4.12d) the trophoblast lineage. In sum, the accessibility of these ETBS-C3-mLA highly accessible sites is unaffected by *Cdx2* knockdown because CDX2 binding here is recruited and/or opportunistic.

'Accessible' ETBS-C3-mLA sites in TSCs are enriched for homeodomain, EOMES and EHF motifs (Figure 4.12b). This indicates that CDX2 binding at these sites may be predominantly driven by binding to its own consensus motif. However, the accessibility of these sites and the expression of *Eomes* and *Elf5* are not affected by *Cdx2* knockdown (Chapter 4.3), so the accessibility of ETBS-C3-mLA 'accessible' sites is probably dependent on EOMES and ELF5 binding rather than CDX2. Functionally, ETBS-C3-mLA accessible peaks are not enriched around genes that are expressed in the trophoblast lineage (Figure 4.12e), nor trophoblast-related mouse phenotypes (data not shown). Instead, the most enriched GO biological processes term is negative regulation of stem cell differentiation (Figure 4.12f).

At 'inaccessible' ETBS-C3-mLA sites, 19 of the top 20 most enriched motifs are homeobox motifs (Figure 4.12b). As discussed in Chapter 3.4, it is not clear whether CDX2 is a pioneer factor, but its binding at 'inaccessible' ETBS-C3-mLA sites suggests that CDX2 can bind condensed chromatin. In addition to being 'inaccessible', these CDX2 binding sites are not enriched around trophoblast-related genes (data not shown) and are instead located around genes associated with regulation of histone modifications and chromatin organisation (Figure 4.12g). In all, these sites are unaffected in *Cdx2* knockdown because they inaccessible and, as a result, the biological relevance of CDX2 binding at these sites is unknown.

### 4.5.2 Histone modifications are distinct at different CDX2 binding subsets

To provide further context to these different CDX2 binding clusters, published histone ChIP-seq data was integrated into analyses (Chuong et al., 2013). Histone marks can be indicative of active or inactive gene function. H3K27ac is an active enhancer marker that is critical for controlling enhancer-regulated GRNs and the production of enhancer RNAs (Creyghton et al., 2010, Rada-Iglesias et al., 2011, Raisner et al., 2018). H3K4me1 also marks enhancers, but not their activity, and is dispensable for enhancer function (Dorighi et al., 2017). The strength of H3K4me3 signal, present at the promoter-TSS of active genes, correlates with the strength of gene expression (Santos-Rosa et al., 2002, Heintzman et al., 2007, Bernstein et al., 2005, Okitsu et al., 2010, Benayoun et al., 2014). Both H3K27me3 and H3K9me3 are repressive markers (Rea et al., 2000). H3K27me3 is enriched on the inactive X chromosome and is responsible for long-term cell type-specific transcriptional repression (Plath et al., 2003, Silva et al., 2003, Lee et al., 2006, Boyer et al., 2006, Ezhkova et al., 2009). In contrast, H3K9me3 marks repeat-rich constitutive heterochromatin and is specifically concentrated at pericentric and centric heterochromatin (Nakayama et al., 2001, Peters et al., 2003).

There are three CDX2 binding site peaksets outlined above in which CDX2 is probably recruited by its own consensus motif; LA, ETBS-C3-mLA accessible, and ETBS-C3-mLA inaccessible. At CDX2 binding sites in which *Cdx2* knockdown causes a loss in chromatin accessibility, there is some enrichment for the generic enhancer marker H3K4me1 but not the active enhancer marker H3K27ac (Figure 4.13). Similarly, most ‘accessible’ ETBS-C3-mLA sites are enriched for the enhancer marker H3K4me1, but are not marked as active (Figure 4.13). In contrast, few ‘inaccessible’ ETBS-C3-mLA CDX2 binding sites show enrichment for both the enhancer markers H3K27ac and H3K4me3 (Figure 4.13).

The homeobox motif was minimally enriched in all other CDX2 binding site clusters (ETBS-C1-mLA, ETBS-C2-mLA and ETBS-C3-mLA ‘highly accessible’), indicating that CDX2 binding here is recruited by other genes and/or is opportunistic because these sites are highly accessible. CDX2 binding sites that are highly accessible in mESCs, ETBS-C1-mLA, are primarily located at promoter-TSS as shown by the high

enrichment of H3K4me3 reads (Figure 4.13). Some CDX2 binding sites that were less accessible in mESC, ETBS-C2-mLA, are found at promoter-TSS. However, ETBS-C2-mLA and ETBS-C3-mLA 'highly accessible' peaksets are mostly active enhancers (Figure 4.13).

Overall, as well as the different enrichment of GO terms and motifs, CDX2 binding site subsets have distinct histone modification profiles, indicating that they have different functions in TSCs.

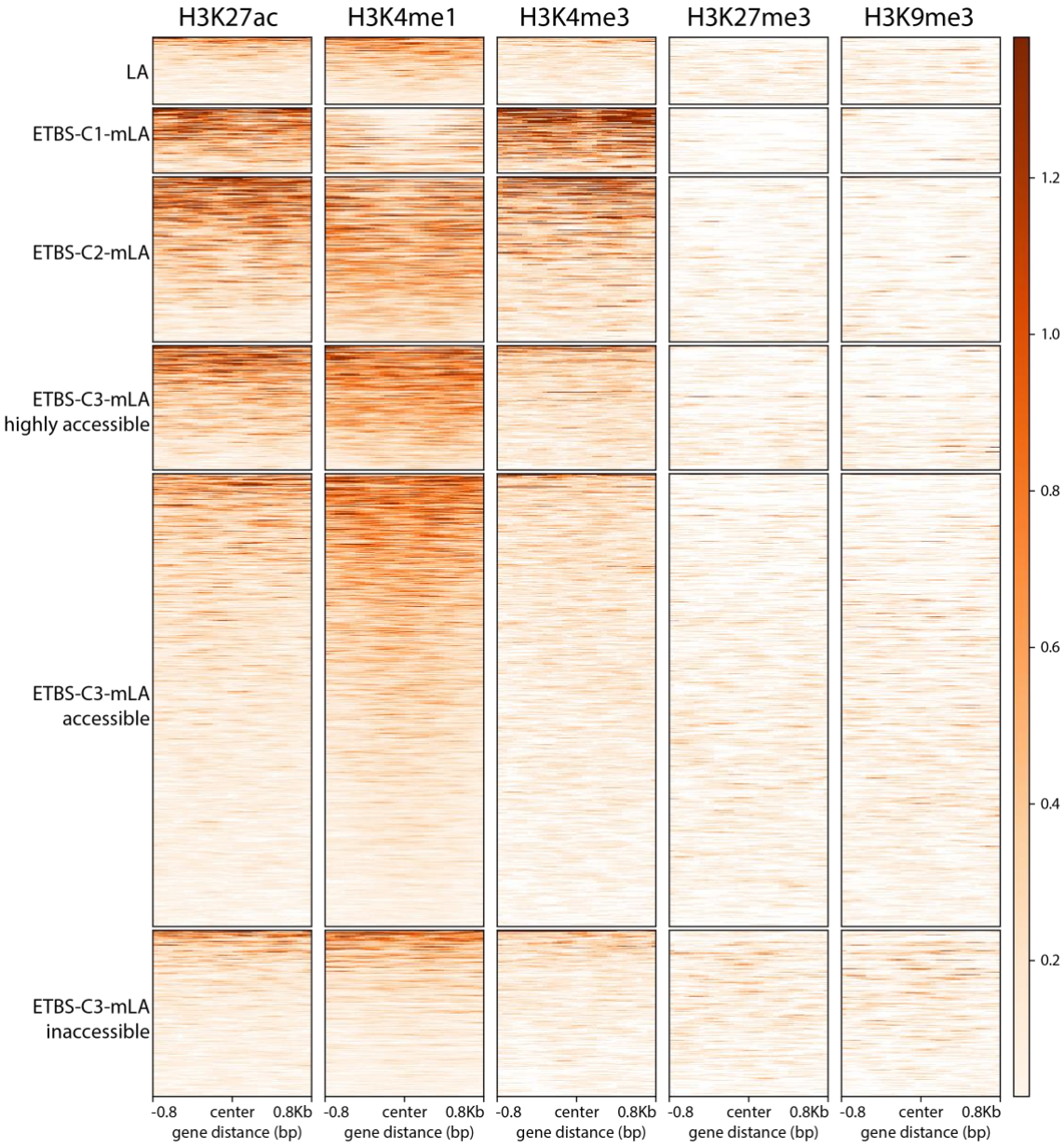



**Figure 4.13: Histone modifications at CDX2 binding site subsets.** Read density heatmap for histone modification ChIP-seq data generated in TSCs at different CDX2 binding site subsets.

### 4.5.3 Tfap2c, gained accessibility and histone modifications

According to motif and footprinting analyses, increased accessibility under *Cdx2* knockdown conditions results from the upregulation of *Tfap2c*. A TSC marker gene, TFAP2C is upregulated during trophoblast differentiation and is the key driver of the global increase accessibility associated with this process (Latos et al., 2015b, Nelson et al., 2017). In agreement with the importance of *Tfap2c* to trophoblast differentiation, sites of gained accessibility are located around genes that are expressed in trophoblast differentiation (Figure 4.9g). To ask whether sites of gained accessibility in *Cdx2* knockdown are bound by TFAP2C in TSCs or one day after differentiation by growth factor withdrawal (DD1), I compared my gained accessibility sites to published TFAP2C ChIP-seq datasets (Latos et al., 2015b).

This analysis revealed that 52% of gained accessibility sites overlap with called TFAP2C binding sites in either TSCs, DD1 or both (Figure 4.14a). One possible explanation for the low overlap between these peaksets is that these sites are inaccessible in wild-type TSCs. To determine how TFAP2C binding correlates with accessibility in wild-type TSCs and *Cdx2* knockdown TSCs, read density heatmaps were generated (Figure 4.14b). Most of these sites are poorly accessible in wild-type TSCs conditions but sites still show some enrichment for TFAP2C binding in one or both libraries (Figure 4.14b). In all, this suggests that TFAP2C does bind sites of gained accessibility, but there is no difference in its binding at these sites between TSCs and one day after growth factor withdrawal to differentiate cells.

Next, I compared how similar TFAP2C binding sites are between TSCs and DD1 to assess how much TFAP2C binding changes over a short period of time. Although these libraries were only generated one day apart, TFAP2C is redistributed rapidly across the genome during differentiation, with 23.88% and 17.17% of TSC and DD1 peaks unique, respectively (Figure 4.14a). However, TFAP2C binding at differential binding sites between TSCs and DD1 is weak compared to its binding at shared sites in both conditions (Figure 4.14c). In summary, TFAP2C binding is redistributed to different loci across the genome rapidly under differentiation conditions but the sites that TFAP2C binds most strongly to do not change between TSCs and one day after growth factor withdrawal.

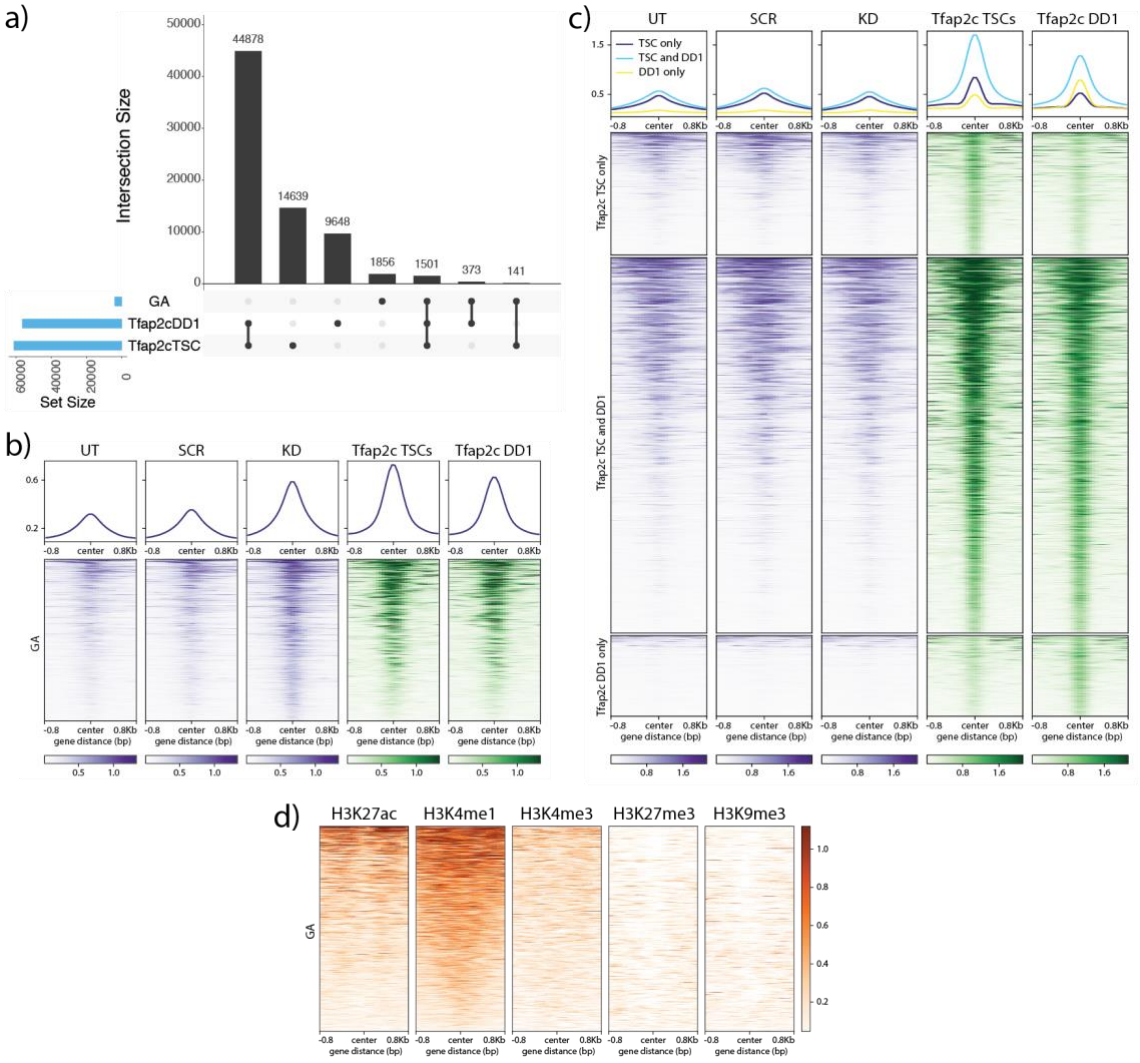



**Figure 4.14: Integration of published TFAP2C and histone marker ChIP-seq data at gained accessibility sites in *Cdx2* knockdown.** a) UpSetR plot to show the intersection of TFAP2C ChIP-seq generated in TSCs and one day after initiation of differentiation by growth factor withdrawal (DD1) with sites of gained accessibility in *Cdx2* knockdown (GA). b) Heatmap to show the read density of UT, SCR and KD ATAC-seq (purple) and TFAP2C TSC and DD1 ChIP-seq (green) libraries at sites of gained accessibility (GA). c) Read density heatmap showing chromatin accessibility (ATAC-seq, purple) and TFAP2C binding (ChIP-seq, green) at all TFAP2C binding sites. Peaks were separated into unique (TSC only, DD1 only) and overlapping (TSC and DD1) peaksets by intersecting peaks with HOMER. d) Read density heatmap of histone modification marker ChIP-seq data generated in TSCs at different GA sites in *Cdx2* knockdown.

The sites that gain accessibility in *Cdx2* knockdown are different from gained TFAP2C binding sites in differentiation induced by growth factor withdrawal (Figure 4.14a). To understand the significance of these sites, I next asked what histone modifications are enriched at sites of gained accessibility in *Cdx2* knockdown (Figure 4.14d). This revealed that these chromatin regions are strongly marked as enhancers by H3K4me1. As expected given that these are mostly inaccessible, few of these enhancers are marked as active (H3K27ac, Figure 4.14d). Further work is required to determine if these sites subsequently become active in *Cdx2* knockdown.

### 4.5.4 Observed transcriptional and chromatin accessibility changes in Cdx2 knockdown poorly correlate

Most GA sites are highly enriched for H3K4me1, a general enhancer marker (Figure 4.14d). To see if these regions may be enhancers, I asked whether changes in chromatin accessibility observed in *Cdx2* knockdown resulted in dysregulated genes. For this, the closest lost (LA) or gained accessibility (GA) sites to the transcription start site (TSS) of every differentially expressed gene was determined (Figure 4.15a). This revealed that the closest chromatin region showing any differential accessibility in *Cdx2* knockdown was ≥50kb or away for 56.8% and 54.8% of all upregulated and downregulated genes, respectively. These sites, therefore, either represent distal enhancers or are unrelated to the genes dysregulated by *Cdx2* knockdown.

To determine if there is a trend that links the direction of chromatin accessibility and transcriptional changes in *Cdx2* knockdown, the distance of the closest LA and GA site to each upregulated and downregulated gene TSS, respectively, was assessed (Figure 4.15b-c). GA sites are more commonly located within ±50kb of the TSS of upregulated genes than LA sites (38.3% versus 8.27% LA sites) (Figure 4.15b). Similarly, LA sites are more commonly found within ±50kb of the TSS of downregulated genes than GA sites (32.7% vs 17.9% total closest peak) (Figure 4.15c). Whether these are biologically relevant observations is unclear. Overall, the relationship between chromatin accessibility and transcriptional changes is poor.

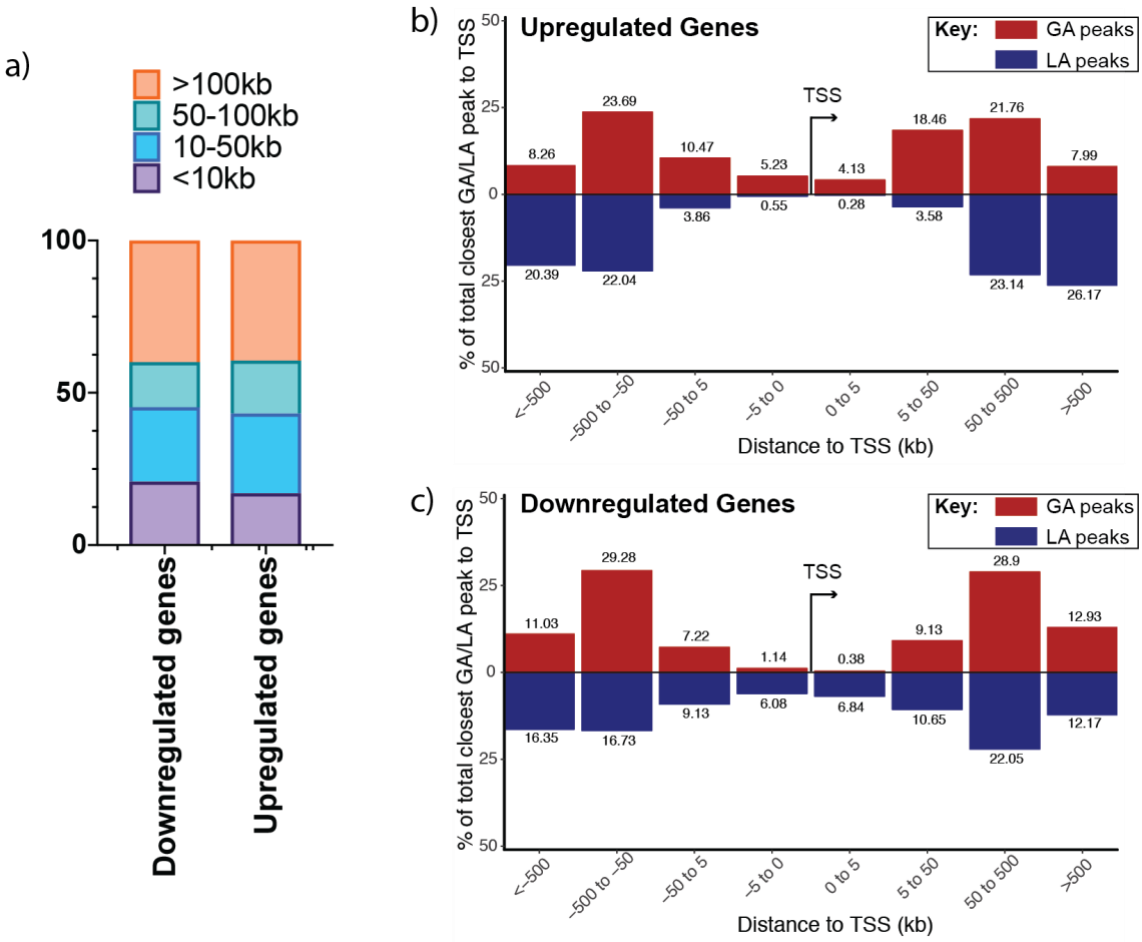



**Figure 4.15: Proximity of differentially accessible chromatin regions in *Cdx2* knockdown to the TSS of dysregulated genes.** a) Bar chart showing the closest differential accessibility site to the TSS of all down- and upregulated genes in *Cdx2* knockdown. b) Bar chart to show the closest gained accessibility (GA, red) and lost accessibility (LA, blue) sites to the TSS of every upregulated gene. c) Bar chart to show the closest GA (red) and LA (blue) sites to the TSS of every downregulated gene

## 4.6 Summary

Short-term *Cdx2* knockdown perturbs both the transcriptome and the chromatin landscape of TSCs, with almost 5% of genes changing in their expression levels and 6192 chromatin regions changing in accessibility. Fold changes of differentially expressed genes are small, perhaps because knockdown was only of the order of 53-62%. Nevertheless, my results provide new insights into *Cdx2* function in the trophoblast fate and new information about the gene regulatory network that underpins TSC maintenance.

Upregulated genes account for 58% of all dysregulated genes and are enriched for GO terms related to cell adhesion and other differentiation pathways. Genes that are downregulated in *Cdx2* knockdown are enriched for pathways related to the immune response and apoptosis. This is in direct contrast to the effect of *Cdx2* knockdown *in vivo*, where trophectodermal cells show a significant increase in apoptosis (Strumpf et al., 2005). Therefore, *Cdx2* knockout in the trophectoderm *in vivo* and *Cdx2* knockdown in TSCs *in vitro* have different transcriptional consequences. This provides further evidence that *Cdx2* plays distinct roles during establishment of the trophoblast lineage and in maintaining the trophoblast stem cell status.

Beyond this, the exact role of *Cdx2* is difficult to ascertain given the poor correlation between transcriptional and chromatin landscape perturbations. Whilst sites that lose accessibility after short-term *Cdx2* knockdown appear to be caused by a direct loss of CDX2 binding, they are low in number and their function is unclear. GO analysis reveals that these sites are localised around genes that are important in cell junctions. Whilst abnormal junctions are observed in *Cdx2*-null trophectoderm *in vivo* (Strumpf et al., 2005), these genes were not shown to be differentially expressed in RNA-seq analyses. Alternatively, sites of lost accessibility in *Cdx2* knockdown are enriched around genes that are expressed in multiple *Cdx2*-related lineages, perhaps indicating that they represent core CDX2 binding sites irrespective of the cell type. This would provide an explanation as to why the direct biological function of lost accessibility sites is difficult to elucidate.

Only 6% of CDX2 binding sites in embryo-derived TSCs (ETBS) lose accessibility after short-term Cdx2 knockdown. In Chapter 3.5, I separated CDX2 binding sites into clusters based on their accessibility in mESCs. Highly accessible (ETBS-C1) and accessible (ETBC-C2) binding sites are not enriched for the canonical homeobox motif. Instead, CDX2 binding at these regions is probably opportunistic or recruited by other transcription factors, such as TEAD4 and EOMES, binding to their canonical motifs. This is further substantiated by the previously established direct interaction between CDX2 and both TEAD4 and EOMES (Latos et al., 2015a). Accordingly, only 21.6% of decreased accessibility in *Cdx2* knockdown overlap with peaks in both of these clusters. CDX2 binding sites that are inaccessible in mESCs, ETBS-C3, are highly enriched for the homeobox motif and binding at these regions are amongst the strongest of all sites in TSCs. However, when subdivided further based on their accessibility in TSCs, ETBS-C3 also generates peaksets that are poorly enriched for the homeobox motif.

Therefore, the accessibility of CDX2 binding sites is largely independent of CDX2 binding itself, and is instead dependent on binding of other transcription factors. Other than at those sites that showed reduced accessibility in *Cdx2* knockdown, consensus homeobox motifs are most enriched in CDX2 binding sites at less accessible or inaccessible chromatin regions in TSCs. As these genomic regions remain inaccessible over successive passages, they do not represent active pioneer transcription factor (TF) sites. However, it is possible that these sites represent passive pioneer TF sites that facilitate binding of other unknown factor(s) that are not expressed within stable TSCs. If these are indeed passive pioneer TF binding sites, what factor(s) they recruit is undetermined and not represented by motif analysis.

Given the small number of genes differentially expressed after *Cdx2* knockdown, determining where *Cdx2* fits into the TSC GRN is difficult. The expression of most TSC marker genes, including *Eomes* and *Elf5*, are unaffected by short-term *Cdx2* knockdown, revealing that they are not direct downstream targets of *Cdx2*. In contrast, TSC marker genes *Gata3* and *Tfap2c* are upregulated in *Cdx2* knockdown. *Gata3* is upstream of *Cdx2* in established TSCs, regulating *Cdx2* expression using an enhancer within intron 1 of *Cdx2* (Home et al., 2009). The observed upregulation of *Gata3* in *Cdx2* knockdown may be indicative of a feedback loop between *Gata3*

and *Cdx2* that controls both of their expression levels. However, the enhancer region within *Cdx2* intron 1 shows no increase in accessibility in *Cdx2* knockdown cells over wild-type controls (data not shown). As *Gata3* is also upregulated in differentiation (Ralston et al., 2010), its upregulation may also indicate that *Cdx2* knockdown cells are differentiating as predicted by GO analysis on sites of increased accessibility.

Sites of gained accessibility, accounting for almost two thirds of all changes to the chromatin landscape in *Cdx2* knockdown, were attributed to TFAP2C binding. As with *Gata3*, *Tfap2c* controls *Cdx2* expression via an enhancer within *Cdx2* intron 1 (Cao et al., 2015). Therefore, its upregulation after *Cdx2* knockdown may result from a feedback loop between these their expression. *Tfap2c* expression levels are also critical to balance the TSC GRN. When TSC are differentiated *in vitro* by growth factor withdrawal, the associated upregulation of TFAP2C drives differentiation and a global increase chromatin accessibility (Latos et al., 2015b, Nelson et al., 2017). In agreement with the importance of *Tfap2c* to TSC differentiation, sites of increased accessibility in *Cdx2* knockdown are found around genes associated with placental differentiation, especially differentiation towards trophoblast giant cells (TGCs). Although GA sites are also enriched for the generic enhancer marker H3K4me1, these regions are over 50kb from most genes that are dysregulated in *Cdx2* knockdown indicating that these are distal enhancers or have minimal immediate effect on the transcriptome.

The TGC marker *Hand1* is also significantly upregulated in *Cdx2* knockdown. It is well established that *Hand1* overexpression alone is capable of inducing differentiation of TSCs into TGCs, albeit in only 30% of targeted cells (Scott et al., 2000). Together with the gene ontology analysis of GA sites, the upregulation of three genes, *Tfap2c*, *Hand1* and *Gata3*, indicates that long-term *Cdx2* knockdown may drive differentiation towards the TGC lineage.

# Chapter 5. Long-term *Cdx2* knockdown drives trophoblast stem cells to differentiate homogeneously into trophoblast giant cells

Transient *Cdx2* knockdown in TSCs perturbs gene expression and alters chromatin accessibility. At some CDX2 binding sites, loss of *Cdx2* causes decreased accessibility, but elsewhere in the genome there is an increase in accessibility driven by the upregulation of *Tfap2c*. Genes whose expression is perturbed by loss of *Cdx2*, as well as those associated with a decrease in accessibility, are not enriched for trophoblast-related terms in gene ontology analyses. However, sites of increased accessibility are associated with genes that are involved in trophoblast differentiation, particularly towards trophoblast giant cells (TGCs).

*In vitro*, removal of the growth factors that normally maintain cells as trophoblast stem cells causes them to differentiate into functional trophoblast derivates (Tanaka et al., 1998). This process mimics *in vivo* differentiation, as cells move away from the niche environment of the extraembryonic ectoderm into the ectoplacental cone. Analysis of microarray, RNA-seq, ATAC-seq and ChIP-seq datasets derived from normal TSCs and those subject to growth-factor withdrawal (Nelson et al., 2017, Latos et al., 2015b, Chuong et al., 2013) confirm that *Tfap2c* and *Gata3* are upregulated as TSCs differentiate (Latos et al., 2015b, Ralston et al., 2010). At the same time, *Cdx2* and other TSC markers genes, including *Eomes* and *Esrrb*, are downregulated (Donnison et al., 2015, Latos et al., 2015b). Recent work has shown that TSC differentiation is accompanied by increases in chromatin accessibility that are driven by TFAP2C (Nelson et al., 2017). However, *Tfap2c* drives *Cdx2* expression during trophectoderm establishment, as does *Tfap2c* overexpression in mESCs (Cao et al., 2015, Kuckenberg et al., 2010). Given that *Cdx2* knockdown causes a global increase in TFAP2C-asscociated chromatin accessibility around genes associated with differentiation, the relationship between *Tfap2c* and *Cdx2* in established cells is likely different from that in trophoblast establishment.

In this chapter, I use genetic selection to maintain *Cdx2* knockdown for over 48 hours, and use these cells to explore the long-term effects of lack of CDX2 on TSC

maintenance and potency. As predicted by GO analyses, sustained reduction of *Cdx2* expression proved to drive homogeneous differentiation into TGCs. In contrast, differentiation by growth factor withdrawal, the standard method to differentiate TSCs *in vitro*, generated a heterogeneous population of trophoblast derivatives. To understand why growth factor withdrawal and *Cdx2* knockdown have such different outcomes, I made use of published growth factor withdrawal ATAC-seq and RNA-seq datasets and compared them to my INX datasets. Given the observed long-term *Cdx2* knockdown phenotype, I focused enhancer activity assays on sites of increased chromatin accessibility in the vicinity of the TGC marker gene *Hand1*.

## 5.1 Maintained *Cdx2* knockdown differs from differentiation by growth factor withdrawal and drives parietal trophoblast giant cell formation

Loss of *Cdx2* in transdifferentiated TSC-like cells causes them to differentiate (Niwa et al., 2005, Kuckenberg et al., 2010). In INX TSCs, *Cdx2* returns to wild-type levels without selection and cell morphology comes to resemble that of normal TSCs (data not shown). Therefore, changes to the transcriptional and chromatin landscape after transient *Cdx2* knockdown are not sufficient to change TSC potency.

To obtain effective long-term knockdown of *Cdx2* I used RNAi vectors carrying an mCherry-IRES-puromycin (mCIP) cassette, allowing me to select for transfected cells that continue to express the mCIP cassette. The XYO and Rossant-GFP cell lines were resistant to puromycin and could not be used in these experiments, but experiments on INX TSCs showed that 32-hour treatment with 0.7 µg/mL puromycin was sufficient to kill all untransfected cells (data not shown).

INX TSCs were therefore transfected with KD or SCR constructs and puromycin added 48 hours post-transfection to select for transfected cells. These cells were also compared with INX TSCs induced to differentiate in a heterogeneous manner by growth-factor withdrawal (-GF) (see 2.1.1.4). Puromycin was removed after 32 hours of treatment, and all cells analysed six days after transfection.

By five days after transfection (1.5 days post-puromycin removal), there were clear differences in the phenotypic appearances of SCR, KD and -GF cells. At six days, SCR cells formed wild-type colonies, with only sporadic differentiation perhaps caused by the stress of puromycin selection (Figure 5.1a). In contrast, KD cells resembled TGCs; they were larger and more irregularly shaped than SCR cells with prominent nuclei (Figure 5.1a). A few phenotypically normal TSC colonies were present in KD cells at seven days, potentially due to silencing of the shRNA and continued expression from the antibiotic resistance cassette (data not shown). After growth factor withdrawal, cultures appeared heterogeneous, with some cells resembling TGCs (Figure 5.1a).

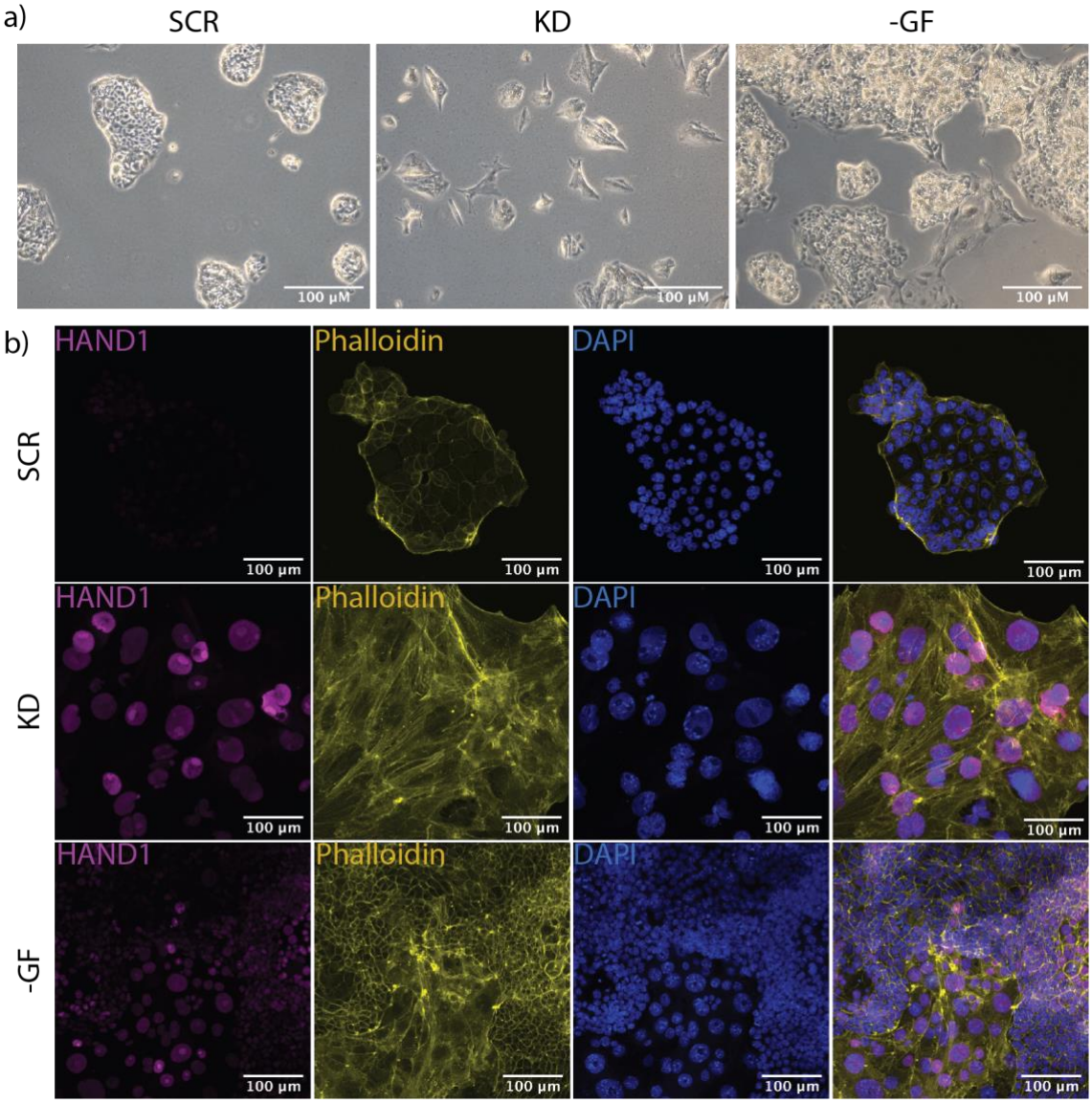



**Figure 5.1: Long-term *Cdx2* knockdown drives homogenous differentiation into TGCs.** a) Brightfield images taken of SCR and KD six days post-transfection and six days post-growth factor (-GF) withdrawal. At 48-hours post-transfection, cells positively transfected with SCR and KD vectors were selected using a 32-hour treatment with puromycin. b) Immunofluorescence staining of HAND1, actin fibres (Phalloidin) and DAPI in SCR, KD and -GF cells six days post-transfection/growth factor withdrawal.

TGCs have large nuclei, prominent actin stress fibres, and express HAND1 (Parast et al., 2001). KD cells exhibit these characteristics in a homogeneous manner, while SCR cells fail to express HAND1 and actin is localised at the cell membrane. Growth factors withdrawal caused a subset of cells to exhibit the characteristics of TGCs (Figure 5.1b) (Tanaka et al., 1998).

Previous work has shown that *Elf5*, *Eomes* and *Esrrb* are downregulated after five days of *Cdx2* knockdown (Latos et al., 2015a). qRT-PCR demonstrated that the same is true in KD and -GF conditions compared with SCR controls (Figure 5.2).

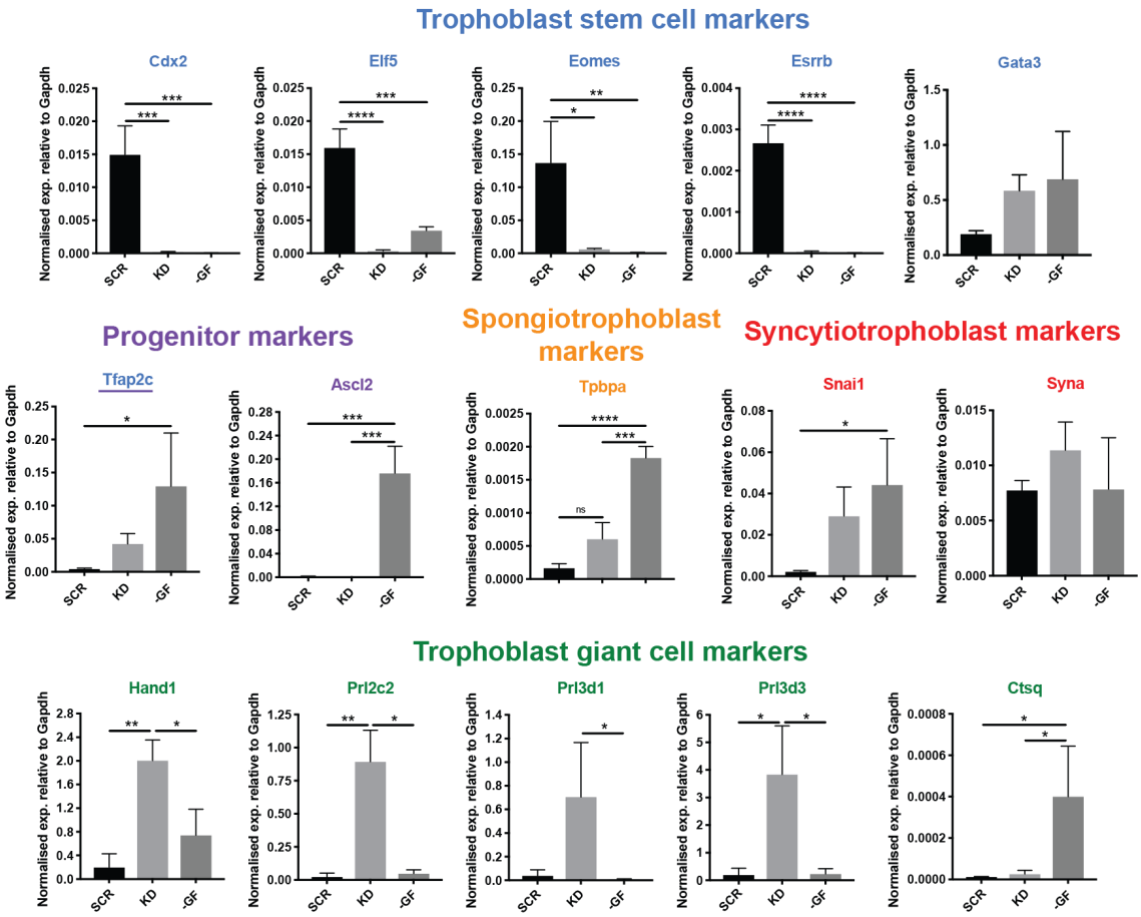



**Figure 5.2: Differentiation induced by *Cdx2* knockdown and growth factor withdrawal are distinct.** qRT-PCR analysis of trophoblast lineage markers in SCR, KD and -GF samples 6 days post-transfection or growth factor withdrawal. Those trophoblast lineages assayed are trophoblast stem cells, progenitors, spongiotrophoblast, syncytiotrophoblast and trophoblast giant cell. Data are normalised to *Gapdh*. n=3 biological replicates per cell line. Significance determined by one-way ANOVA statistical test: * = <0.0332, ** = <0.0021, *** = <0.0002, **** = <0.0001. Error bars display SD.

Other markers of differentiation were also studied. *Gata3* and *Tfap2c*, both TSC markers, are upregulated during growth factor withdrawal differentiation (Latos et al., 2015b, Ralston et al., 2010) and the same was true in KD and -GF cells (Figure 5.2). Interestingly, the ectoplacental cone marker *Ascl2* is upregulated in -GF cells but not in KD cells, suggesting that KD cells do not follow a typical TSC differentiation route (Figure 5.2). *Tpbpa*, a marker of the ectoplacental cone and spongiotrophoblast fates, is upregulated in -GF conditions compared with SCR and KD (Figure 5.2). Similarly, *Snai1*, a marker of the syncytiotrophoblast lineage, was significantly

upregulated only in -GF cells. Unexpectedly, *Syna*, another syncytiotrophoblast marker, was expressed at low levels under all three conditions (Figure 5.2).

I also analysed markers specific for trophoblast giant cell sub-types. The transcription factor *Hand1* is expressed in all four TGC sub-types (Simmons et al., 2007). *Prl3d1* (*Pl1*) is exclusively found in parietal TGCs (p-TGCs). *Ctsq* is expressed in sinusoidal TGCs (s-TGCs) and channel TGCs (Ch-TGCs) (Simmons et al., 2007, Rawn et al., 2015). *Prl3d3* (*Plig*) is expressed in both p-TGC and Ch-TGCs subtypes (Rawn et al., 2015). *Prl2c2* (*Plf*) is expressed in spiral-artery associated TGCs (SpA-TGCs), in canal TGCs (c-TGCs), s-TGCs and p-TGCs.

Expression of *Hand1* proved to be upregulated in *Cdx2* KD cells, as was expression of *Prl3d1*, *Prl2c2* and *Prl3d3*. None of these markers was significantly upregulated in -GF cells compared with SCR, perhaps because SCR cells underwent spontaneous differentiation. The s-TGC and Ch-TGC marker *Ctsq* was only upregulated in -GF samples (Figure 5.2). These results indicate that long-term *Cdx2* knockdown drives cells to differentiate into p-TGCs.

## 5.2 Distinct chromatin landscape changes in *Cdx2* knock-down and differentiation by growth factor withdrawal

In Chapter 4, GO analysis correctly predicted that long-term *Cdx2* knockdown would drive trophoblast stem cells to differentiate into trophoblast giant cells (Chapter 4, Figure 4.9g). This is an indirect effect, driven by an increase in expression and binding of TFAP2C. Previous analyses of the effect of growth factor withdrawal on TSCs indicate that heterogeneous differentiation is similarly accompanied by a global increase in chromatin accessibility, driven by TFAP2C (Nelson et al., 2017). What, then, underlies the differences in TSC differentiation following long-term *Cdx2* knockdown and growth factor withdrawal?

One possibility is that TFAP2C binds differently under the two conditions. To test this idea, I compared the INX ATAC-seq libraries to those generated in the Robertson Lab (RL): one made in TSCs (herein named 'TS RL') and another made from cells two days after growth factor withdrawal (herein named '-GF RL') (Nelson et al., 2017). These datasets were available as processed data files aligned to the mm10 genome, the same genome used for INX libraries, from the NCBI GEO depository (GSE94694) (Nelson et al., 2017) (see 2.5.1, Table 2.2).

There are two potential reasons for differences between RL and INX libraries: biologically relevant differences related to different the cell lines and/or condition used, and biologically irrelevant differences resulting from the analysis pipelines. If biologically irrelevant differences are present between these libraries, this would compromise any future analysis. To address the possibility of biologically irrelevant differences between INX and RL libraries, regions of open chromatin in 'TS RL', '-GF RL' and all INX TSCs ('merged all INX') were compared to see how similar they were (Figure 5.3a). In total, 84.81% of open chromatin in 'TS RL' libraries and 73.52% of '-GF RL' were also open in 'merged all INX' (Figure 5.3a). Although the profiles of RL and INX peaks were different in height and shape, their positions were similar across the genome (Figure 5.3b-c) (Thorvaldsdottir et al., 2013). Therefore, neither the cell line nor the analysis pipeline changed the overall chromatin landscape and TFAP2C driven differences between these libraries can be directly compared.

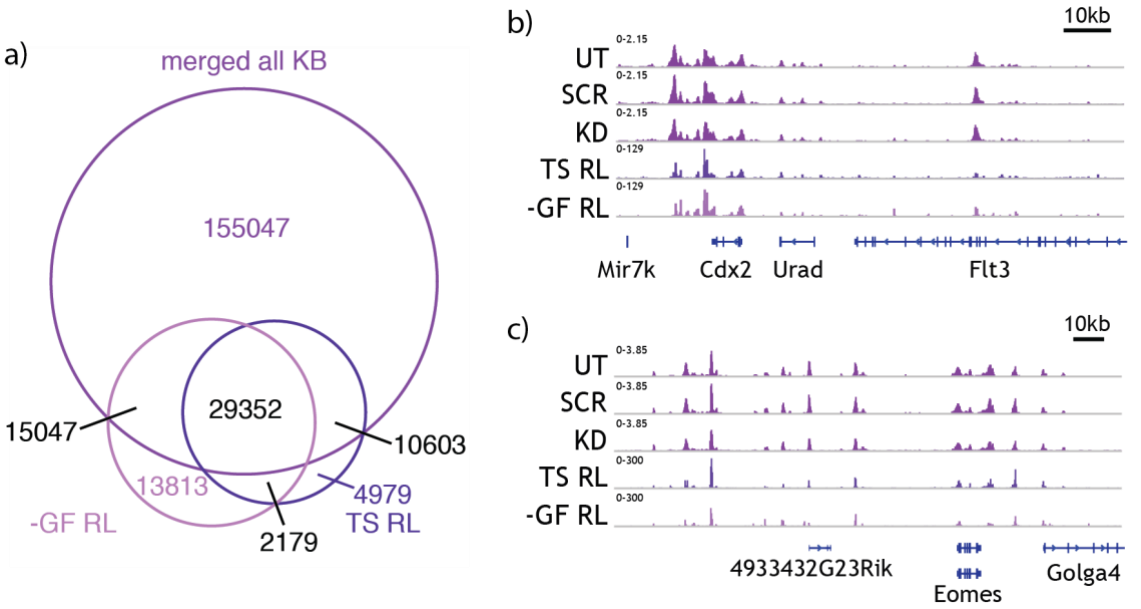



**Figure 5.3: 'TS RL', '-GF RL' and INX ATAC-seq libraries are similar despite the use of different cell lines and analysis pipelines.** a) Venn diagram displaying the overlap of the total accessible regions called for 'TS RL', '-GF RL' and 'merged all KB'. 'Merged all KB' represent all chromatin regions called as accessible in at least one the UT, SCR NEG, SCR or KD ATAC-seq libraries. The discrepancy in peak number between 'TS RL' (57,019) and 'merged all INX' (210,088) are perhaps due to a difference in the number of libraries or the stringency of analysis itself. b-c) IGV tracks to display the UT, SCR, KD, 'TS RL' and '-GF RL' libraries around two core TSC marker genes: *Cdx2* and *Eomes*. As 'TS RL' and '-GF RL' were downloaded as pre-analysed datasets, these are displayed on a different scale to INX libraries.

In total, 7378 and 3879 regions of chromatin were shown to gain accessibility after two days of growth factor withdrawal (d2DARs) (Nelson et al., 2017) or *Cdx2* knockdown (GA, Chapter 4.3), respectively. To compare these peaksets, I found the closest d2DARs peak to every GA peak (Figure 5.4a). Although 12.76% of GA peaks have a d2DAR peak within 100bp, the closest d2DAR peak to 76.54% of GA sites are over 10kb away. Therefore, sites of gain accessibility in *Cdx2* knockdown (GA) are distinct from those that result from growth-factor removal mediated differentiation (d2DARs).

GA and d2DAR peaks were next intersected to determine which sites were shared or unique to *Cdx2* knockdown and growth factor withdrawal conditions. However, GA (507.5bp ± 195.1bp) and d2DARs (193.9bp ± 97.3bp) peaks varied in size because of the different analysis pipelines used. To overcome this, all peaks were made to be 700bp total around their centre point, making them similar in size to the largest GA peaks. When these were intersected, peaks were classified as overlapping (515), GA only (3330) and d2DARs only (6522) (Figure 5.4b). Whilst both *Cdx2* knockdown

(KD) and growth factor withdrawal ('-GF RL') conditions showed increased accessibility over TSCs at overlapping sites, only KD or '-GF RL' increased in accessibility at GA only and d2DARs only sites, respectively (Figure 5.4c). Differences in accessibility were also confirmed locally (Figure 5.4d-e). There is a shared peak at the *Plac1* transcription start site (TSS) (orange box, Figure 5.4d), a gene upregulated after transient *Cdx2* knockdown (Chapter 4.3, Figure 4.4c), as well as multiple d2DARs peaks (black box, Figure 5.4d). In contrast, the *Tfap2c* locus is surrounded by unique sites (Figure 5.4e). Gained accessibility is higher under growth factor withdrawal conditions than *Cdx2* knockdown conditions, both globally (Figure 5.4c) and locally (Figure 5.4d-e). Therefore, these peaksets accurately reflect overlapping and unique sites of gained accessibility between these conditions.

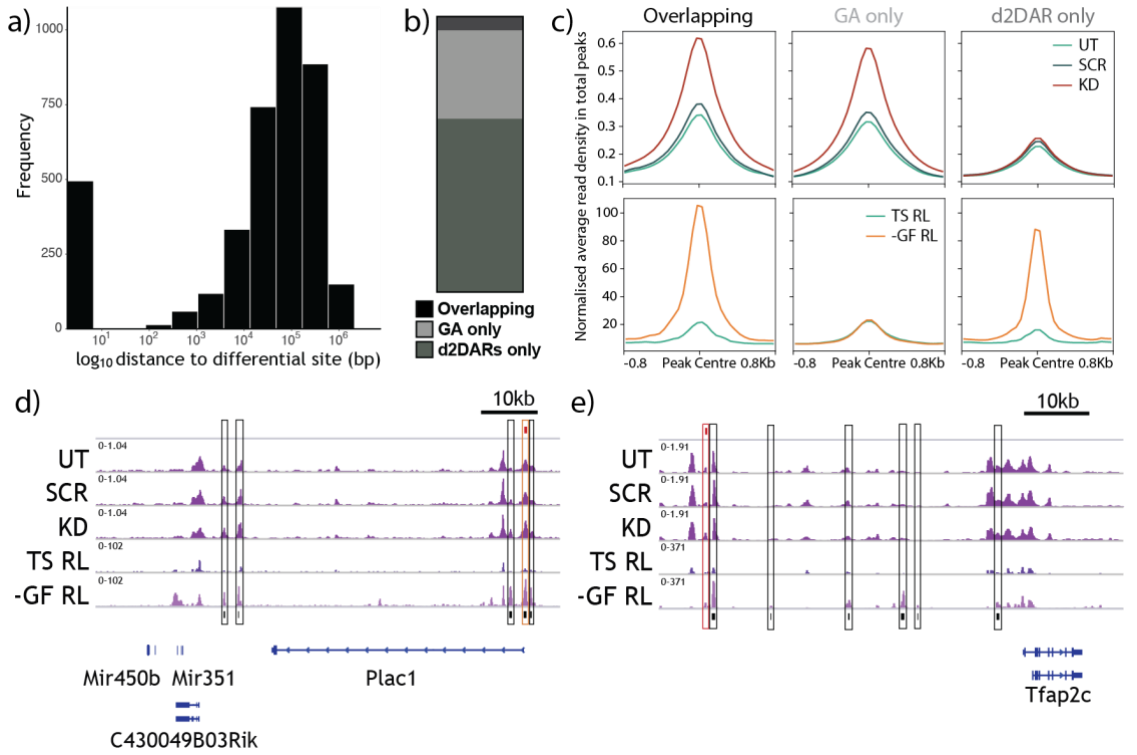



**Figure 5.4: Sites of gained accessibility are distinct between short-term *Cdx2* knockdown and differentiation induced by growth factor withdrawal.** a) Histogram displaying the closest d2DARs peaks to every GA site. b) The centre of every GA and d2DAR peak was determined, extended by 350bp on either side and intersected using HOMER (Heinz et al., 2010). This chart summaries the proportion of overlapping, GA only and d2DARs only peaks. c) The average read density of UT, SCR and KD or 'TS RL' and '-GF RL' ATAC-seq data at overlapping, GA only and d2DAR only peaksets. d-e) IGV track displaying UT, SCR, KD, 'TS RL' and '-GF RL' ATAC-seq data around the *Plac1* and *Tfap2c* locus. Black boxes denote d2DAR only peaks, red boxes represent GA only peaks and orange boxes denote overlapping peaks.

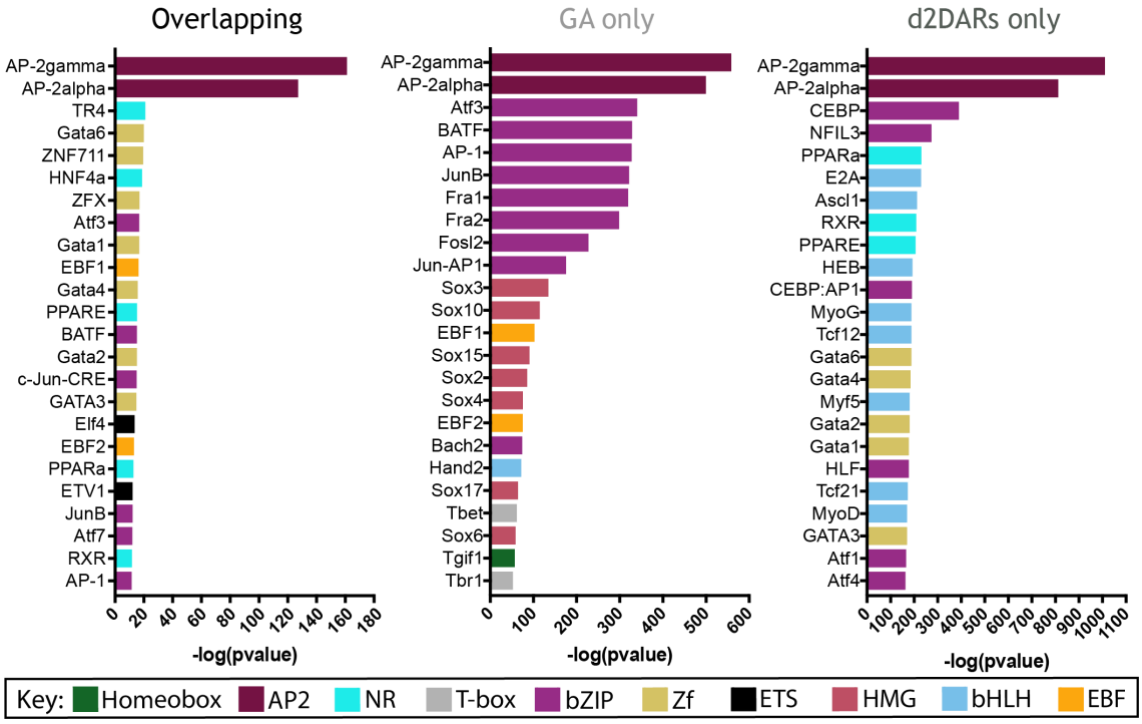



**Figure 5.5: Differential TFAP2C binding in *Cdx2* knockdown and growth factor withdrawal induced differentiation might be caused by different co-factors.** Bar charts displaying the 24 most statistically significant enriched motifs in overlapping (left), GA only (middle) and d2DARs only (right) peaksets as determined using HOMER.

Next, I re-confirmed that TFAP2C is the main driver of gained accessibility at shared and unique sites, with AP-2γ and AP-2α the two most enriched motifs in all peaksets (Figure 5.5). Beyond this, however, enriched motifs in GA only and d2DARs only peaksets are very different. Whilst GA only peaks are enriched for AP-1 transcription factor components and SOX motifs, d2DARS are enriched for GATA, C/EBP, PPAR and E-box bHLH, such as ASCL1, motifs (Figure 5.5). Intriguingly, transcription factor family members for some of these motifs have previously been associated with different trophoblast derivates. Whilst *Sox15* is known to drive TGC formation, *Pparb* and *Ppard* are required to maintain the ectoplacental cone and spongiotrophoblast (Yamada et al., 2006, Nadra et al., 2006, Wang et al., 2007).

## 5.3 Genes perturbed in *Cdx2* knockdown and differentiation by growth factor withdrawal overlap

The differential motif enrichment between GA only and d2DAR only peaks may be driven by differential expression of transcription factor family members between conditions. To test this idea, published RNA-seq libraries from TSCs and two growth factor withdrawal time points, day 1 (DD1) and day 3 (DD3), were incorporated into analyses (Latos et al., 2015b). These libraries were re-analysed using the same pipeline as for INX libraries.

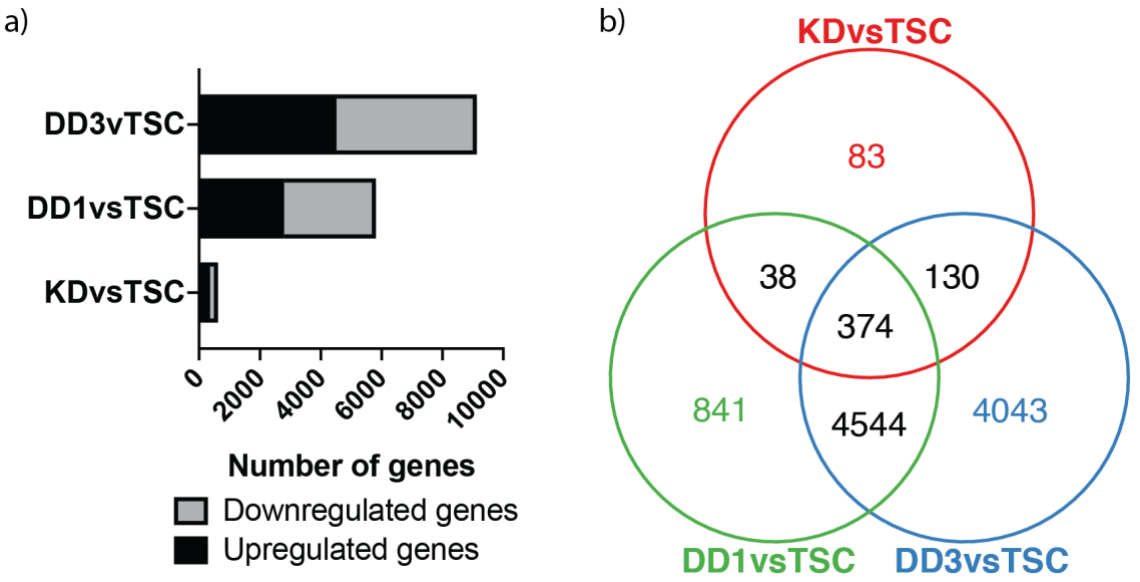



**Figure 5.6: Those genes perturbed by *Cdx2* knockdown highly correlate with those affected when TSCs are differentiated by growth factor withdrawal.** a) Bar charts summarising the number of genes significantly up- and downregulated by *Cdx2* knockdown (KD), one day after growth factor withdrawal (DD1) or three days after growth factor withdrawal (DD3), over TSCs. b) Venn diagram summarising the number differentially expressed genes between KDvsTSC, DD1vsTSC and DD3vsTSC that overlapped.

Despite the short timeframes, 5821 and 9133 genes were differentially expressed one day or three days after growth factor withdrawal, respectively (Figure 5.6a). Although fewer genes are affected by transient *Cdx2* knockdown (626, Chapter 4.3), 86.74% of these overlap with those genes differentially expressed after growth factor withdrawal (Figure 5.6b). Of these overlapping genes, 56.35% are differentially expressed in the same direction (cluster 1 and 2, Figure 5.7a) and 30.94% in opposite directions (clusters 3 and 4, Figure 5.7a) at one or both growth factor withdrawal time points. No GO analysis was performed as this would have provided little insight into the potential functional relevance of such a small number of genes.

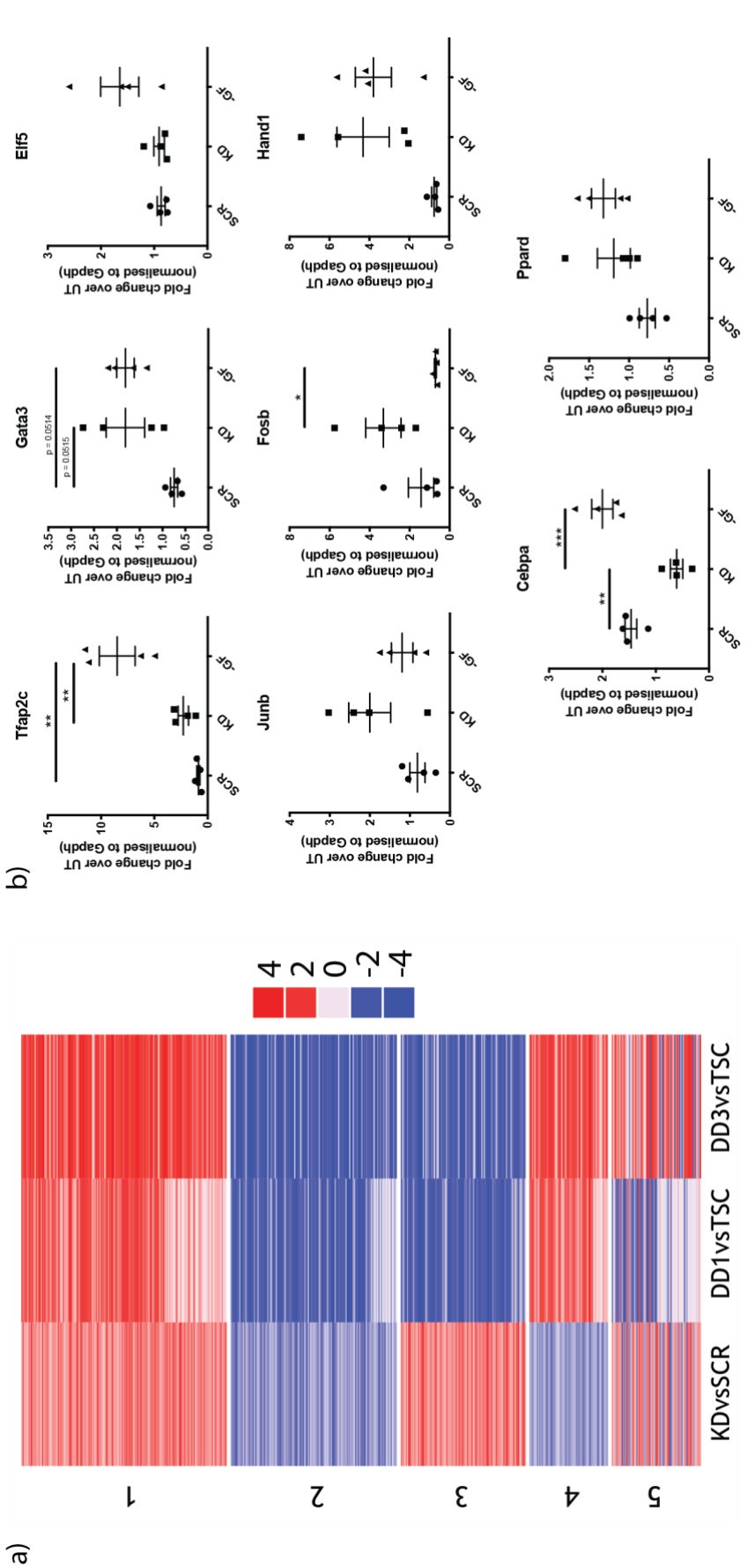



**Figure 5.7: The directionality of overlapping genes that are differentially expressed in *Cdx2* knockdown and differentiation by growth factor withdrawal is not always conserved.** a) Those genes differentially expressed in *Cdx2* knockdown (KD) that overlapped with genes differentially expressed in DD1 and/or DD3 were separated into artificial clusters. For clusters 1-4, DD1 and DD3 had to be differentially expressed in the same direction in both DD1 and DD3 or be differentially expressed in one and non-changing in the other. Clusters 1 and 2 represent upregulated (1) or downregulated (2) genes in KD and DD1 and/or DD3. Clusters 3 and 4 represent genes differentially expressed in opposite directions between KD and DD1/DD3. Cluster 5 represents genes that are also differentially expressed between DD1 and DD3 in opposite directions. b) qRT-PCR analysis of differentially expressed genes under KD and/or DD1/DD3 conditions that are related to trophoblast stem cells or whose families are enriched at sites of gained accessibility in KD or -GF conditions. UT, SCR, KD and -GF samples were collected from concurrent experiments in biological quadruplicate (n=4). Graphs displayed as fold change of gene expression normalised to *Gapdh* over UT samples. Significance determined by one-way ANOVA: * p≤0.0332, ** p≤0.0021, *** p≤0.0002. Error bars display SEM.

qRT-PCR revealed that *Tfap2c* is significantly upregulated in -GF over both the SCR and KD conditions, but is less so in KD over SCR (Figure 5.7b). Whilst TSC marker genes *Gata3 and Elf5* are upregulated after short-term growth factor withdrawal (Latos et al., 2015b, Ralston et al., 2010), only *Gata3* is after *Cdx2* knockdown (Figure 5.7b).

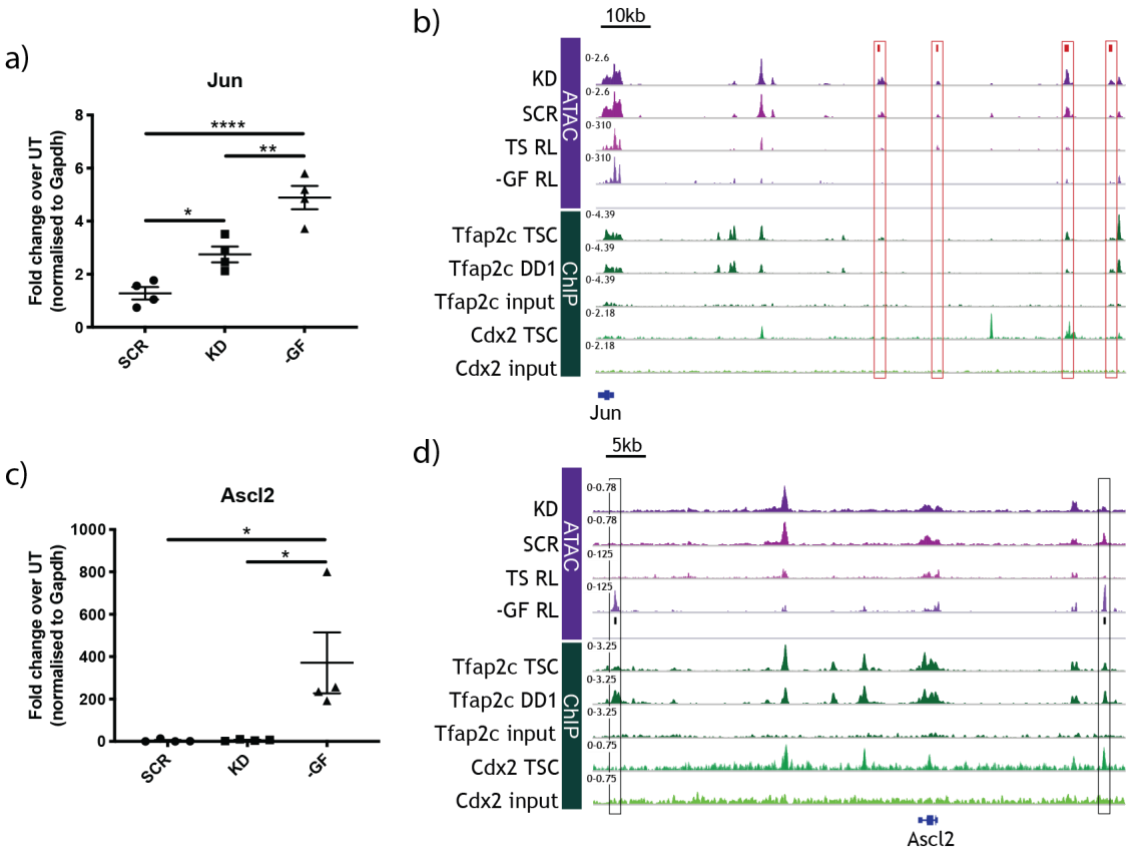



**Figure 5.8: *Jun* and *Ascl2* are differentially expressed and show differential sites of gained accessibility under *Cdx2* knockdown and differentiation by growth factor withdrawal conditions.** a,c) qRT-PCR analysis of *Jun* (a) and *Ascl2* (c). Graphs displayed as fold change (gene expression normalised to *Gapdh*) over UT samples. Significance determined by one-way ANOVA: * p≤0.032, ** p≤0.0021. Error bars show SEM. n=4 biological replicates. b,d) IGV track displaying KD, SCR, 'TS RL' and '-GF RL' ATAC-seq data (purple), TFAP2C ChIP-seq generated in TSCs and DD1 (dark green), and CDX2 ChIP-seq (light green) generated in TSCs around the *Jun* (b) and *Ascl2* (d) loci.

Multiple AP-1 transcription factor family motifs are enriched in GA only sites, including AP-1, JunB, Fra1, Fra2 and Jun-AP1 (Figure 5.5). qRT-PCR confirmed that AP-1 transcription factor family members *Junb* and *Fosb* are uniquely upregulated in *Cdx2* knockdown (Figure 5.7b). In contrast, *Jun*, is significantly more upregulated after growth factor withdrawal than *Cdx2* knockdown (Figure 5.8a). Despite this, the only changes in accessibility around the *Jun* locus are GA sites (Figure 5.8b). The

HAND2 motif is also specifically enriched in the GA only peakset, but *Hand1* is upregulated in both KD and -GF conditions (Figure 5.7b).

In contrast, d2DAR only peaks are uniquely enriched for a variety of motifs including CEBP, PPARα and ASCL1 (Figure 5.5). *Ppard* shows an upregulated trend in KD and -GF conditions over SCR. In contrast, *Cebpa* is upregulated after growth factor withdrawal but is downregulated after *Cdx2* knockdown (Figure 5.7b). *Ascl2* expression is low in TSCs and remains low in *Cdx2* knockdown, but is significantly upregulated in -GF conditions by more than 200-fold over UT, SCR and KD (Figure 5.8c). There are two d2DARs around the *Ascl2* locus, one upstream and one downstream of the gene, but no differentially accessible regions in *Cdx2* knockdown cells (Figure 5.7d). This may reflect the observed unchanging expression of *Ascl2*.

In all, this suggests that genes whose family motifs are differentially enriched between GA only and d2DAR only peaksets are also differentially expressed between *Cdx2* knockdown and differentiation by growth factor withdrawal conditions.

## 5.4 Long-term *Cdx2* knockdown avoids progression through a progenitor fate

Previous work has shown that cells can differentiate into TGCs *in vivo* without progressing through a *Tpbpa*-positive outer ectoplacental cone (EPC) progenitor fate (Simmons et al., 2007). *Ascl2*, a gene that marks the entire EPC, is not upregulated at either 48-hours (Figure 5.8d) nor six days (Figure 5.2) after *Cdx2* knockdown, suggesting that these cells may not progress through a progenitor fate as they differentiate. *Hand1* and *Ascl2* expression overlaps in the ectoplacental cone where they antagonise one another (Scott et al., 2000). *Ascl2* is essential for maintaining the progenitor fate and, when overexpressed, can prevent differentiation into TGCs. In contrast, *Hand1* overexpression can drive TSCs to differentiate into TGCs, although this is less efficient than by *Cdx2* knockdown (Scott et al., 2000). Unlike *Ascl2*, *Hand1* is upregulated at both the transcript (Figure 5.7b) and protein (Figure 5.9a) level after short-term *Cdx2* knockdown, providing an early indication that *Cdx2* knockdown drives TGC formation.

To determine when morphological difference can first be observed between KD and SCR, live cell imaging was performed immediately after 32-hour puromycin selection. Figure 5.9b shows the same example SCR and KD colonies every 5 hours over a 40-hour time period. In SCR conditions, cells show highly dynamic movement, undergo clear cell divisions and form colonies with characteristic TSC morphology (Figure 5.9b). Whether KD cells undergo cell division after selection is unclear, but they show morphological differences to SCR cells within 5-10 hours of puromycin removal (Figure 5.9b). Giant cell characteristics and endoreduplication are present within 25 hours (Figure 5.9b). In all, this suggests that KD cells, unlike SCR cells, have minimal proliferative capacity and are committed to differentiate into TGCs within 24-hours after they are selected by puromycin treatment.

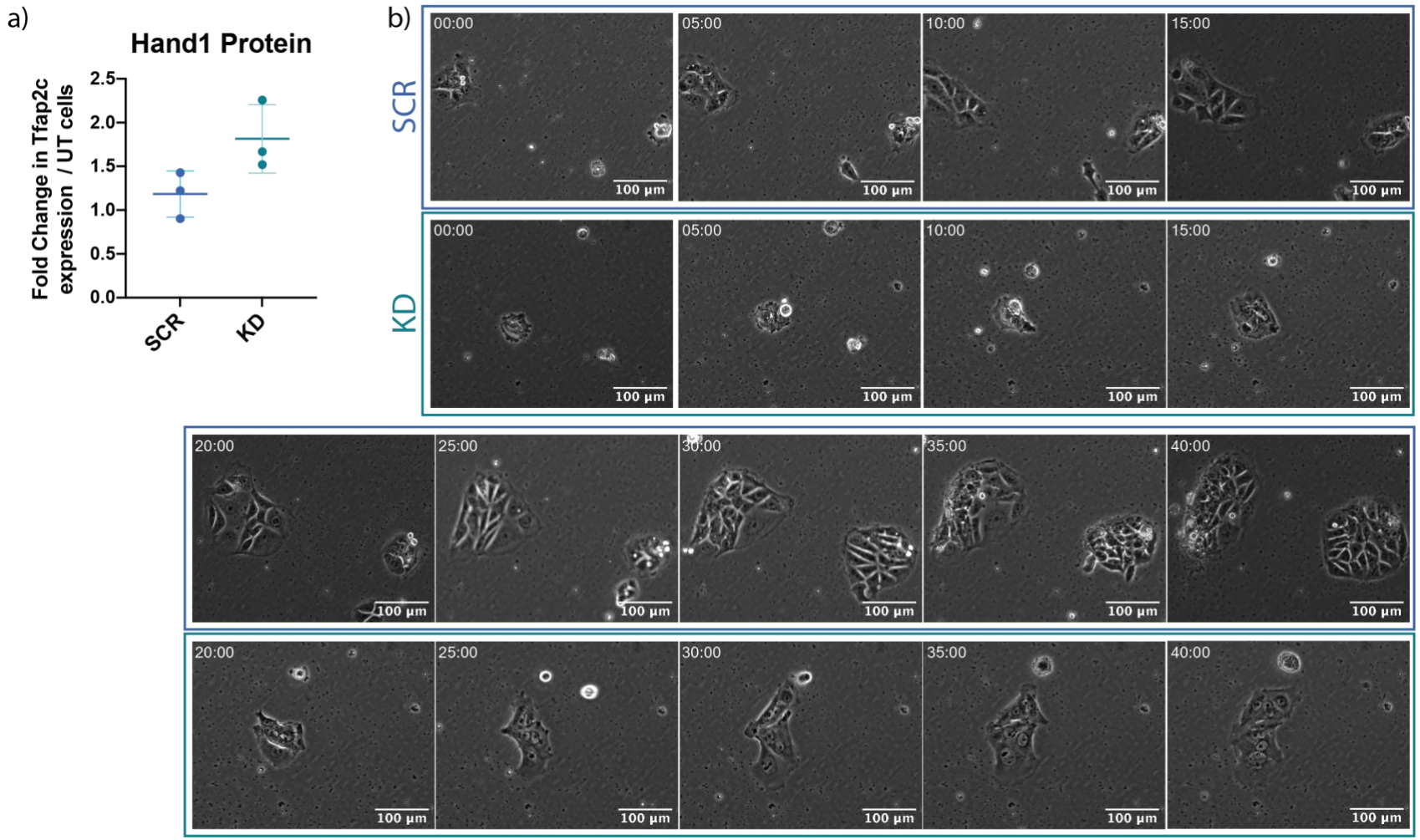



**Figure 5.9: TSCs appear to commit to trophoblast giant cell differentiation within a day after puromycin selection.** a) Fold change in the geometric mean HAND1 protein expression in KD and SCR over UT 48-hours post-transfection as assessed by flow cytometry. Error bars show SD. n=3 biological replicates. T-tests determined that the increase in HAND1 expression in KD over SCR was not significant. b) Stills of live cell imaging on exemplar SCR and KD colonies every 5 hours for 40 hours. Live cell imaging began immediately after 32-hour puromycin selection.

## 5.5 Validation of potential enhancer regions around the *Hand1* locus

The upregulation of *Hand1* after short-term *Cdx2* knockdown is a key indicator that long-term *Cdx2* knockdown drives homogeneous TGC differentiation. There are multiple sites of increased accessibility after short-term *Cdx2* knockdown upstream of the *Hand1* locus that are unaffected by growth factor withdrawal. These genomic loci may, therefore, represent enhancers that are unique to *Cdx2* knockdown. Despite these differences in sites of gained accessibility around its locus, *Hand1* is also upregulated at the transcript level after growth factor withdrawal (Figure 5.7b). As such, the *Hand1* locus represents a model system for determining whether differentially accessible chromatin regions between short-term *Cdx2* knockdown and growth factor withdrawal differentiation conditions are functional.

Whether two loci are likely to interact with one another can be predicted based on how the genome is structured in 3D. Chromosomes themselves occupy distinct and pre-defined nuclear territories (Cremer and Cremer, 2010). Within these territories, the genome is organised into domains known as topologically associated domains (TADs) (Nora et al., 2012, Szabo et al., 2019). As chromatin inside TADs interact more frequently than chromatin in different TADs, gene loci and their enhancer/repressor elements are usually contained in the same TAD domain. Our understanding of the 3D chromatin landscape comes from technological advancements to assay chromatin organisation (reviewed in Sati and Cavalli, 2017).

While circular chromosomal conformation capture (4C) examines how frequently all genomic loci interact with a specific region of interest, Hi-C examines how frequently genomic loci interact with all other genomic loci (Zhao et al., 2006, Belton et al., 2012). Figure 5.10 is a ‘virtual 4C plot’, generated using the 3D Genome Browser, to displays how frequently genomic regions in this window interact with the *Hand1* locus in mESCs (Figure 5.10) (Wang et al., 2018, Bonev et al., 2017). *Hand1*, on the reverse strand, frequently interacts with genomic regions approximately 100kb upstream and 50kb downstream (Figure 5.10). Six sites of increased accessibility in *Cdx2* knockdown are found in the highly interactive upstream region.

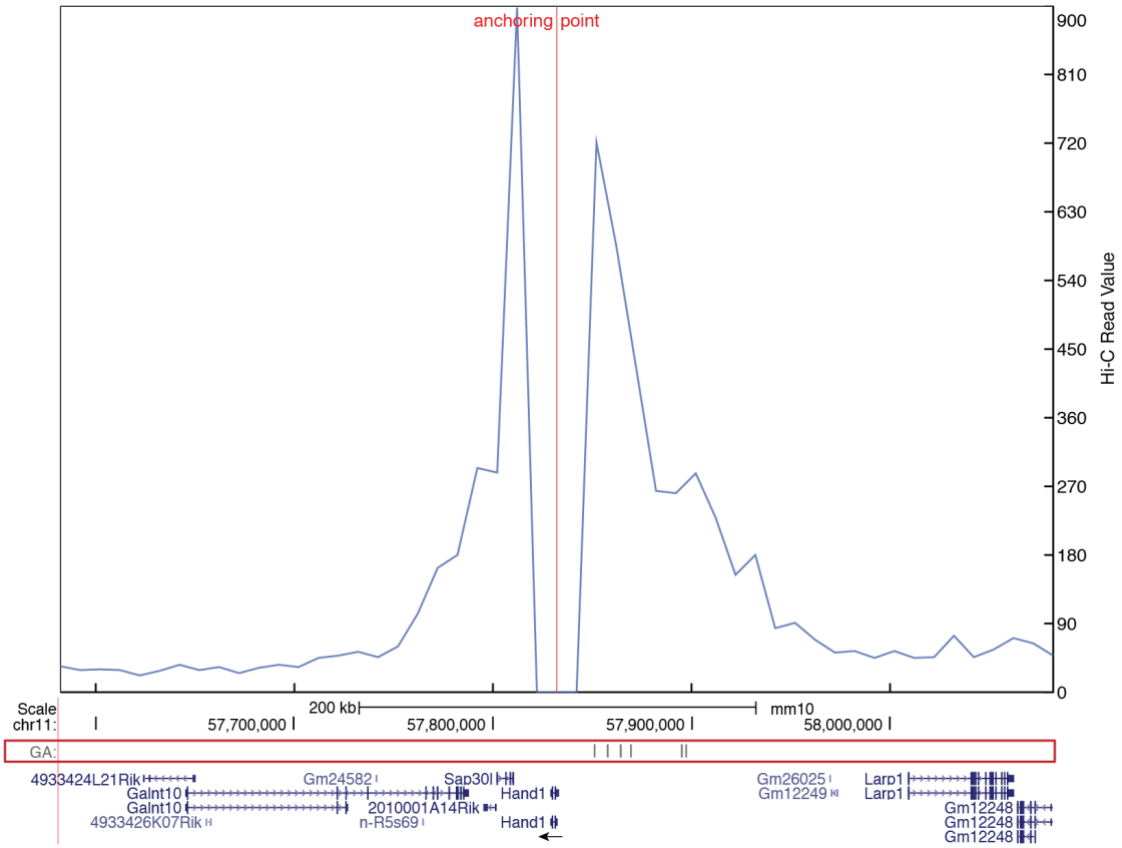



**Figure 5.10: Visualising *Hand1* interactions with surrounding chromatin.** A virtual 4C plot generated by the tool '3D Genome Browser'. This tool enables the most frequent interactions between any locus and the surrounding genome to be visualised from a choice of published Hi-C datasets. This plot displays *Hand1*'s interaction with chromatin 200kb upstream and downstream in mESCs. The higher the Hi-C Read Value, the more frequently two genomic loci interact.

To determine whether these GA sites are enhancer regions for *Hand1*, I performed enhancer validation assays. These were to be tested on the *Hand1* promoter itself. As such, I tested the baseline activity and responsiveness of three *Hand1* 'promoters' of different lengths, containing different promoter elements (Figure 5.11a). The smallest 'promoter', 209bp in length (P1), contains a TATA box but not the Hand1 TSS, whereas the 283bp 'promoter' (P2), contains both. A larger 673bp 'promoter' region (P3) contained a CAAT box and a second TATA box (Figure 5.11a). The activity of each vector was tested in TSCs as these contain all the transcription factors and co-factors that are required to drive the *Hand1* promoter in TSCs.

Each of these promoter constructs (containing Firefly luciferase) was co-transfected with the pRL *Renilla* Luciferase vector to control for transfection efficiency between biological replicates and produce a normalised luciferase readout. Several pGL3 to pRL ratios and acquisition times were tested to determine which were the most

appropriate (Figure 5.11b). All promoter constructs responded similarly, regardless of the condition tested (Figure 5.11b). Therefore, it was decided that 120 ng of pRL *Renilla* Luciferase and a one second acquisition time were to be used for subsequent experiments as this resulted in the highest Firefly:Renila ratios.

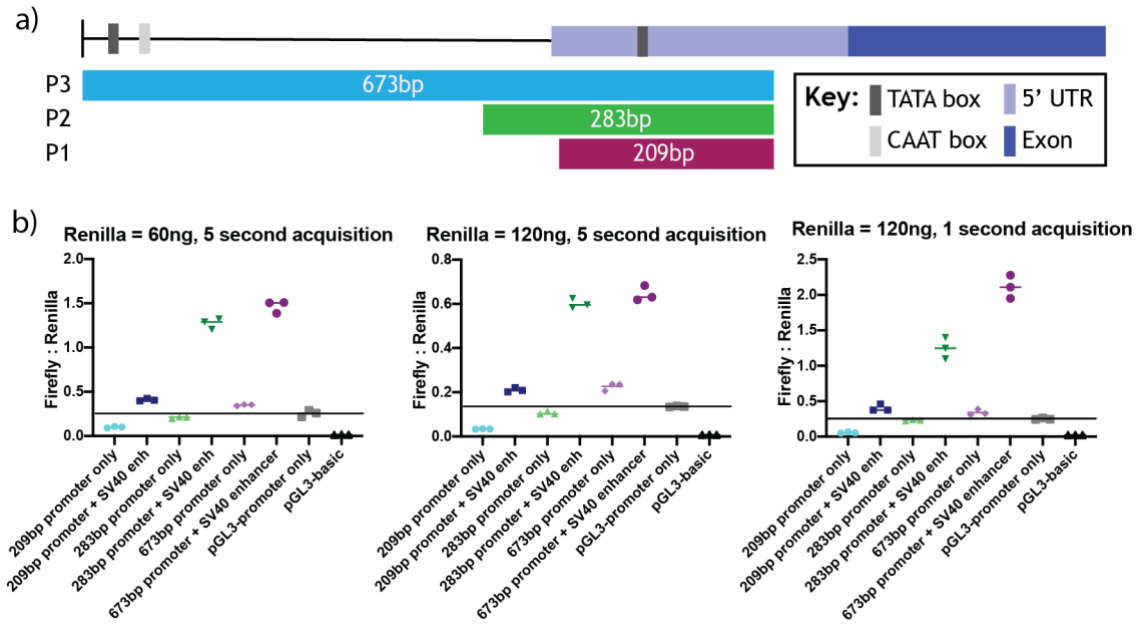



**Figure 5.11: Comparison of different *Hand1* promoters for enhancer validation assays.** a) Schematic representing the different promoter constructs tested for their activity in the dual luciferase assay. b) The baseline activity (promoter only) and responsiveness to the SV40 enhancer was tested for each promoter (P1-3). Different ratios of renilla (pRL) and promoter vectors (firefly) were assayed: either 1:19 (60ng renilla to 1140ng firefly) or 1:9 (120ng renilla to 1080ng firefly). Different acquisition times were also tested: either five seconds or one second.

Whilst the 209bp promoter (P1) responds to the SV40 enhancer, its baseline activity is poor (Figure 5.11b). In contrast, the baseline activity of the 283bp promoter (P2) is similar to the SV40 promoter (pGL3-promoter vector) and increases three- to five-fold when activated by the SV40 enhancer (Figure 5.11b). P3, the largest Hand1 promoter region tested (673bp), is slightly more active than P2 but is similarly responsive to the SV40 enhancer (Figure 5.11b). As the luciferase activity of P2 was most similar in activity to the pGL3-promoter vector, prospective enhancer constructs were cloned onto this vector.

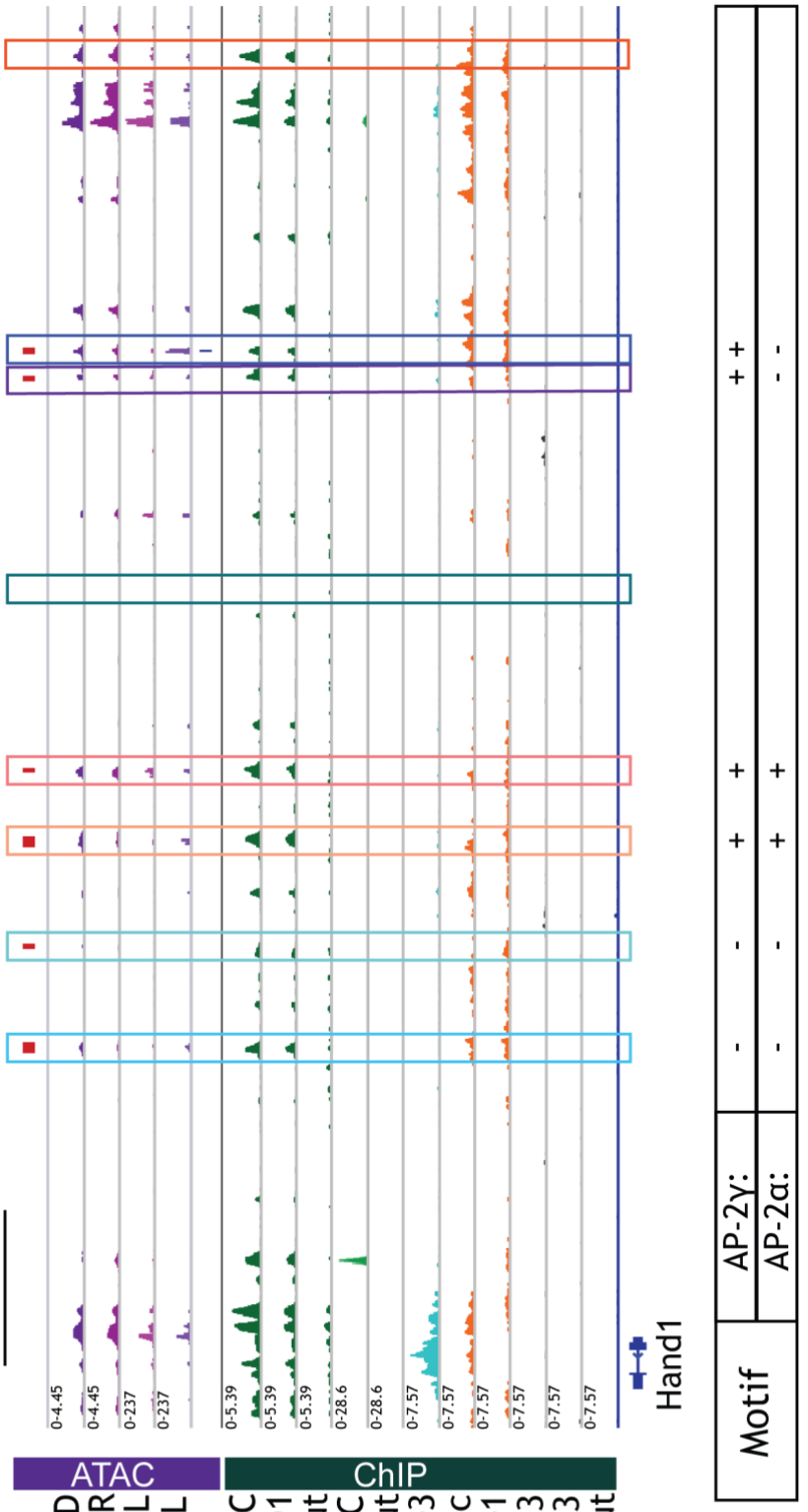



**Figure 5.12: An overview of the chromatin landscape around the *Hand1* locus and those regions assayed for their enhancer activity.** IGV track displaying chromatin accessibility (purple tracks), TFAP2C binding (dark green tracks), CDX2 binding (light green tracks) and histone modification ChIP-seq around the *Hand1* locus. H3K4me3 marks the promoter-TSS (light blue tracks), H3K27ac and H3K4me1 mark enhancers (orange tracks) and H3K9me3 and H3K27me3 are markers of repressed chromatin (black tracks). Those regions shown to gain accessibility in *Cdx2* knockdown (KD) or after two days differentiation by growth factor withdrawal (-GF RL) over TSCs are highlighted by red (GA sites) or dark blue (d2DARs) boxes above or below ATAC-seq tracks (purple). Those regions assayed for their enhancer activity are A-F, C1 and C2. Given that TFAP2C is thought to be the main driver of increased accessibility in *Cdx2* knockdown, whether A-F contain consensus AP-2$\gamma$ or AP-2$\alpha$ motifs are also specified below each region.

Regions that were tested in this enhancer validation assay are outlined in Figure 5.12. Regions A-F all show increased accessibility after short-term *Cdx2* knockdown. Only one of these regions, F, also showed increased in accessibility two days after growth factor withdrawal (Nelson et al., 2017). Two control regions, C1 and C2, were also included in analyses. C1 is inaccessible in TSCs, whereas C2 is bound by TFAP2C and shows minimal difference in accessibility between conditions (Figure 5.12). Whilst regions C-F contain the AP-2$\gamma$ and/or AP-2$\alpha$ consensus motif, neither A nor B contain either. Despite this, TFAP2C binds to regions A-F in both TSCs and one day post-growth factor withdrawal (dark green, TFAP2C ChIP-seq, Figure 5.12).

The most biologically relevant condition to test these enhancers would be after short-term *Cdx2* knockdown in TSCs, but it is unfeasible to co-transfect five vectors. Instead, validation was performed in the presence (+GF) or absence (-GF) of growth factors to see how enhancer constructs to respond to different levels of transcription factors. Routinely, cells are cultured overnight (16 hours) before they are transfected. If TSCs were cultured overnight without growth factors, there was high cell death after transfection (data not shown). To reduce cell death, growth factor withdrawal and transfection were performed at the same time. Although cell death was greater in the -GF conditions than +GF conditions, transfection efficiency was higher in -GF than in +GF (Figure 5.13a). Cells were collected 24-hours after transfection.

Although regions A, C and F have no effect, regions B, D and E significantly upregulate *Hand1* promoter activity in the presence and absence of growth factors (Figure 5.13b-c). Region D is the most active *Hand1* enhancer, but can only increase *Hand1* promoter activity 2.3-fold in the presence of growth factors; significantly less than the 5-fold increase in *Hand1* promoter activity driven by the SV40 enhancer (Figure 5.13b-c). Therefore, regions B, D and E are not strong enhancers for the *Hand1* promoter alone, but may behave synergistically with other *Hand1* enhancers. Surprisingly, both control regions significantly repress the *Hand1* promoter (Figure 5.13b-c). Although this effect from the inaccessible C1 region is difficult to explain, C2 was chosen because it is accessible in TSCs and is bound by TFAP2C (Figure 5.12). TFAP2C binding at C2 appears to decrease after growth factor withdrawal (dark green, DD1 TFAP2C ChIP-seq, Figure 5.12) and may therefore increase *Hand1* expression after growth factor withdrawal indirectly.

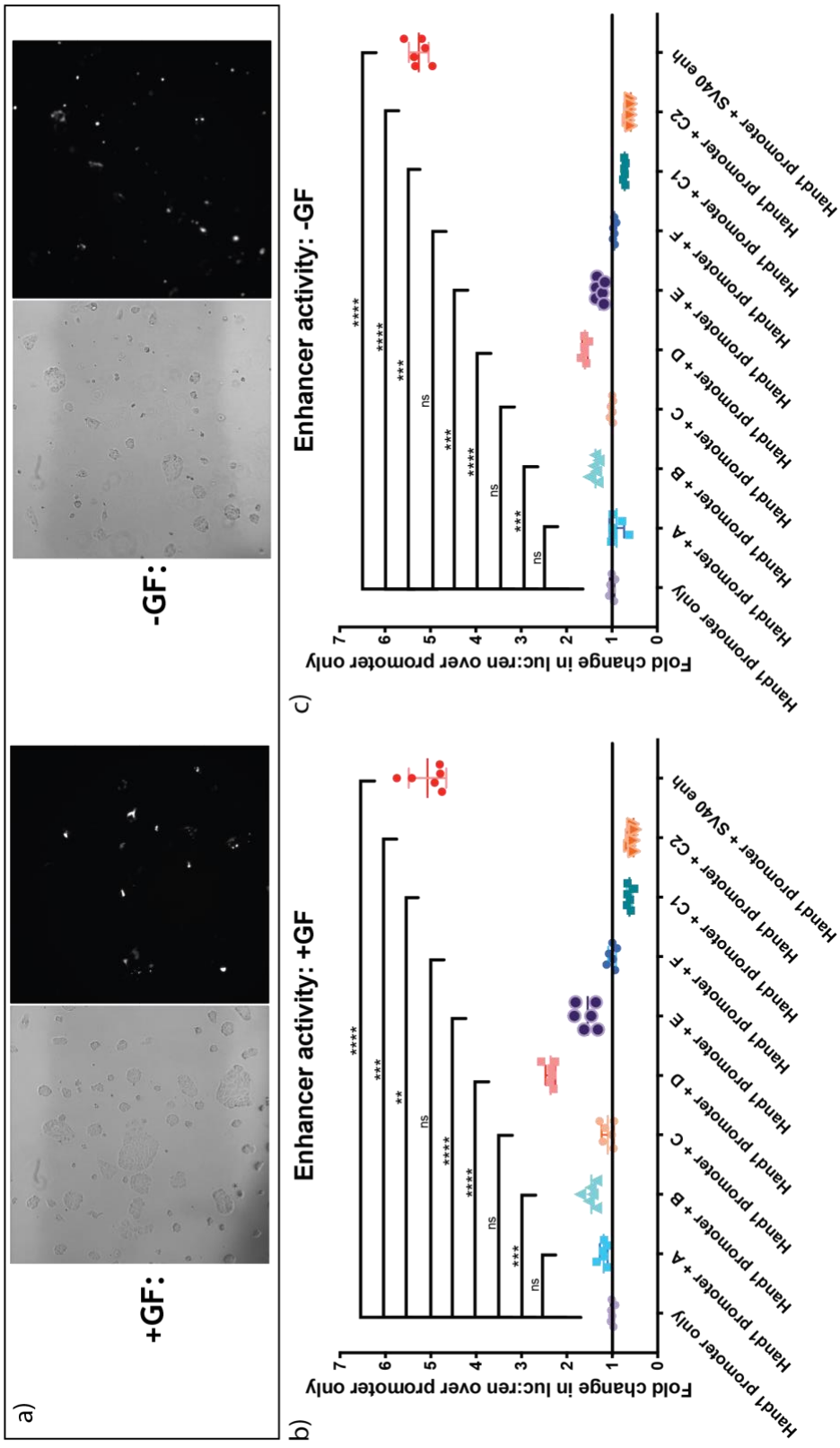



**Figure 5.13: Several sites of gained accessibility in *Cdx2* knockdown are enhancers for the *Hand1* promoter.** a) Images to display cell survival (brightfield) and transfection efficiency (fluorescence) in control wells for transfections performed in the presence (+GF) and absence (-GF) of growth factors. b-c) Fold change in the activity of each promoter + potential enhancer region over the *Hand1* promoter only in dual luciferase assays. Assays were performed in the presence (b) or absence (c) of growth factors. Dual luciferase experiments were performed using biological triplicates run in technical duplicate (3 data points) performed on two separate occasions (n=6 biological replicates). Error bars display SD. Statistical significance was determined by one-way ANOVA: * p≤0.0332, ** p≤0.0021, *** p≤0.0002, **** p≤0.0001

Next, I determined whether the activity of these confirmed enhancers or repressors changed under different growth factor conditions. Both regions D and E were significantly less active under growth factor withdrawal conditions (Figure 5.14a). As the activity of other enhancer-promoter constructs did not change, this decrease in activity is unique to these two enhancers. Unlike region B, both D and E contain the consensus AP-2γ motif (Figure 5.12).

To understand why these enhancers are less efficient after growth factor withdrawal, qRT-PCR was performed on samples collected concurrently with the dual luciferase experiments (Figure 5.14b). *Tfap2c* expression did not change 24 hours after growth factor withdrawal (Figure 5.14b). Therefore, the decline in regions D and E enhancer activity in the absence of growth factors cannot be explained by a change in TFAP2C expression directly. *Cdx2* expression was significantly reduced after growth factor withdrawal and *Ascl2* expression was significantly upregulated. Whilst both *Hand1* and *Gata3* expression increased when growth factors were removed from cultures, this was not statistically significant. In all, this suggests that undetermined factors other than TFAP2C help drive the activity of these enhancers and that these are differentially expressed in the presence or absence of growth factors.

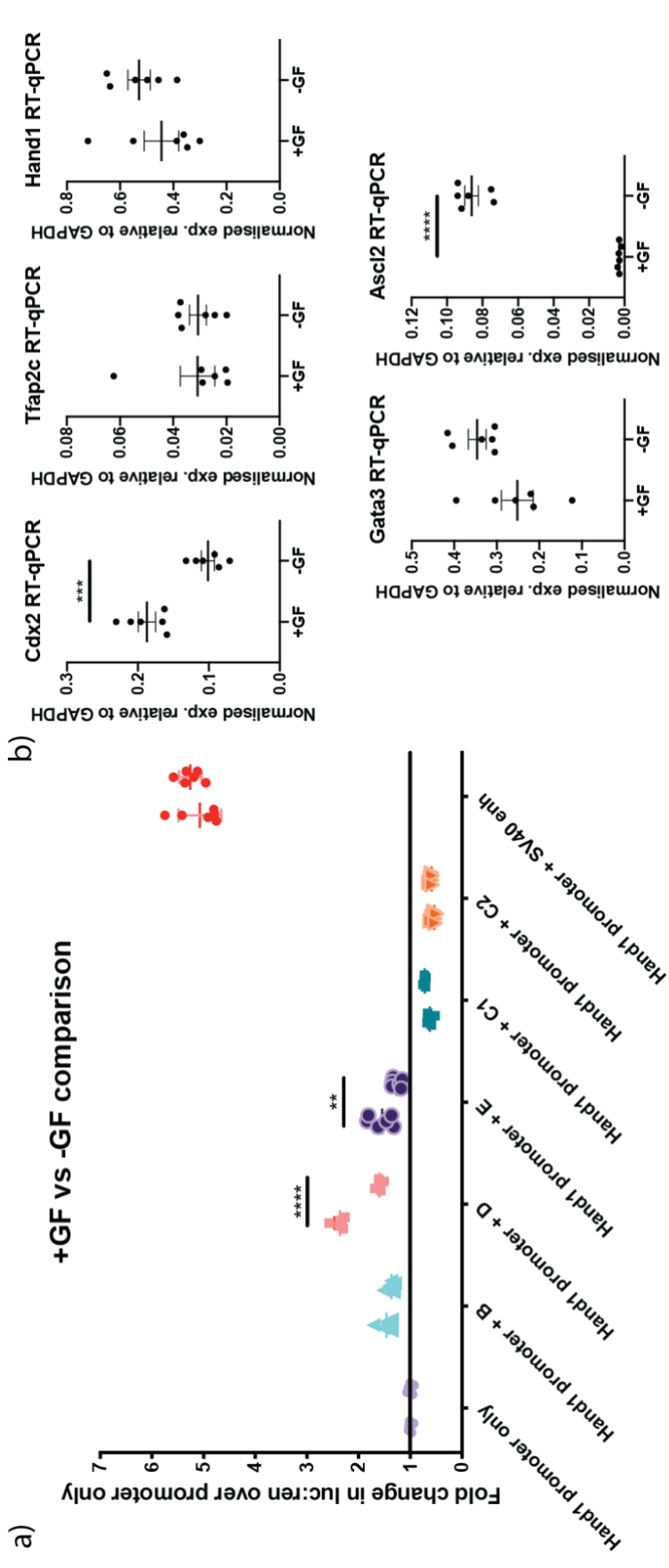



**Figure 5.14: Two *Cdx2* knockdown-specific *Hand1* enhancers in are less active under growth factor withdrawal conditions.** a) Fold change in the activity of each promoter + potential enhancer region over the *Hand1* promoter only in dual luciferase assays. Assays were performed in the presence (left) or absence (right) of growth factors. Dual luciferase experiments were performed as biological triplicates run in technical duplicate (3 data points) performed on two separate occasions (n=6 biological replicates). Error bars display SD. Statistical significance of changes in promoter activity between +GF and -GF conditions was determined by one-way ANOVA: ** p≤0.0021, **** p≤0.0001. b) qRT-PCR for key trophoblast marker genes in control samples collected for +GF and -GF in parallel with dual luciferase assay experiments. Error bars display SEM. Statistical significance between +GF and -GF conditions determined by t-tests: *** p≤0.0002, **** p≤0.0001.

## 5.6 Summary

Sustained *Cdx2* knockdown drives TSCs to homogeneously differentiate into parietal trophoblast giant cells (p-TGCs). *Hand1*, a pan-TGC marker gene, is upregulated in response to short-term *Cdx2* knockdown. Ectopic expression of *Hand1*, even in the presence of TSC maintenance factors, can drive TGCs differentiation, but less efficiently than that observed in *Cdx2* knockdown (Scott et al., 2000, Hughes et al., 2004). Whilst validating my data, I attempted to replicate the findings of Scott et al (2000) and Hughes et al (2004) by overexpressing *Hand1* in the INX cell line. Although I could transiently upregulate *Hand1*, I could neither maintain its expression nor generate a phenotype (data not shown). It is therefore unlikely that sustained overexpression of HAND1 alone is sufficient to drive the homogeneous differentiation observed in *Cdx2* knockdown, at least in the INX cell line.

Instead, the observed homogeneous differentiation is likely driven by upregulation of both *Hand1* and *Tfap2c*. *Tfap2c*, upregulated during differentiation by growth factor withdrawal *in vitro*, drives a global increase in accessibility (Latos et al., 2015b, Nelson et al., 2017). However, these sites of gained accessibility in growth factor withdrawal are distinct from those in *Cdx2* knockdown conditions despite both being attributed to TFAP2C. Further work is required to confirm that TFAP2C binding at these unique sites differs between these conditions. I performed ChIP-qPCR experiments to address this, but they were unsuccessful because of difficulties in obtaining sufficient numbers of cells by cell sorting (data not shown).

Other than AP-2γ and AP-2α motifs, there was no overlap in the motifs enriched at unique sites of gained accessibility in *Cdx2* knockdown and growth factor withdrawal induced differentiation. This indicates that differential TFAP2C binding between *Cdx2* knockdown and growth factor withdrawal conditions might be driven by the use of different co-factors. Indeed, SOX and AP-1 family members motifs were enriched in peaks unique to *Cdx2* knockdown. *Sox15* and various AP-1 family members have previously been implicated in driving trophoblast differentiation, especially towards the TGC lineage (Yamada et al., 2006, Shaulian and Karin, 2002, Renaud et al., 2014, Kubota et al., 2015, Lee et al., 2018). In contrast, PPAR and ASCL1 motifs are only enriched in sites of gained accessibility unique to growth factor withdrawal.

Associated genes *Ascl2*, *Pparb* and *Ppard* are known to be expressed in those progenitor cells derived from the stem cell niched *in vivo*: the ectoplacental cone and spongiotrophoblast (Scott et al., 2000, Nadra et al., 2006, Wang et al., 2007).

Surprisingly, there was a large overlap in the genes perturbed by *Cdx2* knockdown and differentiation by growth factor withdrawal, but the direction in which these genes changed was not always maintained. This disparity in direction of differential gene expression was confirmed for several genes whose motifs were also differentially enriched at TFAP2C binding sites between these two conditions. This re-affirms that differential binding of TFAP2C between *Cdx2* knockdown and growth factor withdrawal are likely influenced by transcriptional differences.

One potential fundamental difference between these TFAP2C-inducing conditions is the unique and rapid upregulation of *Ascl2*, a marker of the progenitor cell-containing ectoplacental cone, after growth factor withdrawal (Cross et al., 1995, Guillemot et al., 1994). The ASCL1 motif is uniquely enriched in sites of increased chromatin accessibility in growth factor withdrawal differentiation. *Ascl2* antagonises *Hand1* expression, maintaining proliferative trophoblast cells and inhibiting giant cell formation (Scott et al., 2000). As such, the lack of *Ascl2* expression may explain why *Cdx2* knockdown cells commit to the TGC fate so rapidly and homogeneously. Indeed, qRT-PCR and live cell imaging experiments suggests that these cells do not progress through a progenitor fate. Previous work shows that *Ascl2* is significantly upregulated within 6-12 hours of growth factor withdrawal, indicating that it is a direct target of the growth factors that maintain TSCs (Donnison et al., 2015). As a result, *Ascl2* upregulation may not occur under *Cdx2* knockdown because these growth factors are present.

Beyond this, the importance of these signalling pathways and the genes under their control was confirmed by the differential activity of two *Cdx2* knockdown-associated *Hand1* enhancers in the presence and absence of exogenously supplied growth factors. Both enhancers contain the AP-2γ and AP-2α consensus motifs and TFAP2C binds to these loci under both TSC maintenance conditions and after one day of differentiation by growth factor withdrawal (Latos et al., 2015b). However, further work is required to confirm the importance of TFAP2C binding to these

enhancer regions. One approach to this would be to mutate the AP-2 consensus binding motifs and then re-examine the efficiency of these enhancers in dual luciferase assays. Indeed, *Tfap2c* expression was the same whether growth factors were exogenously supplied or not, but activity of both enhancers decreased significantly when growth factors were removed. Given this, the underlying cause of this change in responsiveness is likely related to the differences in transcriptional profiles based on the presence and absence of growth factors. The transcription factors that underlie this difference are currently unknown.

In all, this chapter confirms that there are differences in the sites that increase in accessibility between *Cdx2* knockdown and growth factor withdrawal-induced differentiation that are likely critical for their different long-term outcomes. Although the number of sites tested in enhancer validations were small, they confirm that some sites of increased accessibility after *Cdx2* knockdown are functional and responsive to transcriptional changes.

# Chapter 6. Validating the long-term *Cdx2* knockdown phenotype by CRISPR-Cas9 genome editing

I have shown that both the trophoblast stem cell transcriptome and the chromatin landscape are affected by short-term *Cdx2* knockdown. Without selection, *Cdx2* expression reverts to wild-type levels and TSCs are morphologically normal. Yet, if *Cdx2* knockdown is sustained for an extra 32 hours by puromycin selection, INX TSCs are driven homogeneously to differentiate into parietal TGCs (p-TGCs). Thus, although the changes driven by short-term *Cdx2* knockdown are insufficient to commit TSCs to the TGC fate, extension beyond this time, albeit brief, leads to their irreversible commitment to this fate.

Whether these *Cdx2* knockdown cells progress through a progenitor fate as they differentiate into p-TGCs is unclear. All mural trophectoderm and some extraembryonic ectoderm differentiate into p-TGCs without progressing through a progenitor fate (Dickson, 1966, Simmons et al., 2007). *Ascl2*, a key marker of the ectoplacental cone progenitor fate, is not required for primary p-TGC formation (Rossant et al., 1998). Similarly, *Ascl2* is not expressed after 48-hour nor 144-hour *Cdx2* knockdown. Other key trophoblast genes are also affected differently after transient *Cdx2* knockdown; whilst *Elf5*, *Esrrb* and *Eomes* are unaffected by transient *Cdx2* knockdown, their expression is negligible a mere four days later.

The critical events that commit *Cdx2* knockdown TSC to differentiate into TGCs occur during or immediately after puromycin selection. Limited material is available during this time period and selection itself is a source of exogenous stress (Moran et al., 2009). Furthermore, both the XYO and Rossant-GFP cell lines characterised in Chapter 3.1 are puromycin resistant, preventing the validation of the *Cdx2* knockdown phenotype in other TSC lines. Therefore, although informative, this knockdown approach is not appropriate for studying these intermediate stages of TGC commitment.

To address this, I decided to knock out *Cdx2* function in both the INX and Rossant-GFP lines. Recent advancements in CRISPR-Cas9 genome editing approaches

allow specific genomic loci to be targeted using guide RNAs (gRNAs) and a human coding optimised version of Cas9 (reviewed in (Doudna and Charpentier, 2014)). In cell culture systems, targeted cells can be maintained as polyclonal or clonal populations if mutations are not lethal. However, the phenotype of gene knockout and knockdown studies are not necessarily the same (De Souza et al., 2006, Daude et al., 2012). Such differences may arise from genetic compensation in knockout approaches (Rossi et al., 2015, El-Brolosy et al., 2019) or off-target side effects in knockdown approached, including the upregulation of the innate immune response (Gentsch et al., 2018).

In this chapter, I use CRISPR-Cas9 genome editing to generate polyclonal and clonal *Cdx2* mutant TSCs. Using these *Cdx2* mutant TSCs, I validated the importance of *Cdx2* in maintaining the TSC fate by preventing spontaneous TGC formation. Subsequent analysis of clonal *Cdx2* mutant cells provided further understanding of the TSC gene regulatory network.

## 6.1 CRISPR design and efficacy optimisation

Long-term *Cdx2* knockdown drives TSCs to terminally differentiate into p-TGCs (Chapter 5.1) and *Cdx2*-null TSCs cannot be maintained in culture (Niwa et al., 2005). This indicates that it should not be possible to use conventional CRISPR-Cas9 editing approaches to generate clonal populations. Instead, polyclonal *Cdx2*-null TSCs were generated and maintained only in the short term. To achieve this, mutations had to abolish CDX2 function (through frameshifts or large deletions) and efficacy had to be high. By using wild-type Cas9 and no repair template, I aimed to increase the indel rate by favouring non-homologous end-joining (NHEJ) (Yang et al., 2013). Multiple guide RNAs (gRNAs) were designed to target the *Cdx2* homeobox domain and boost disruption of CDX2 function in mutants (Figure 6.1a) (Ran et al., 2013).

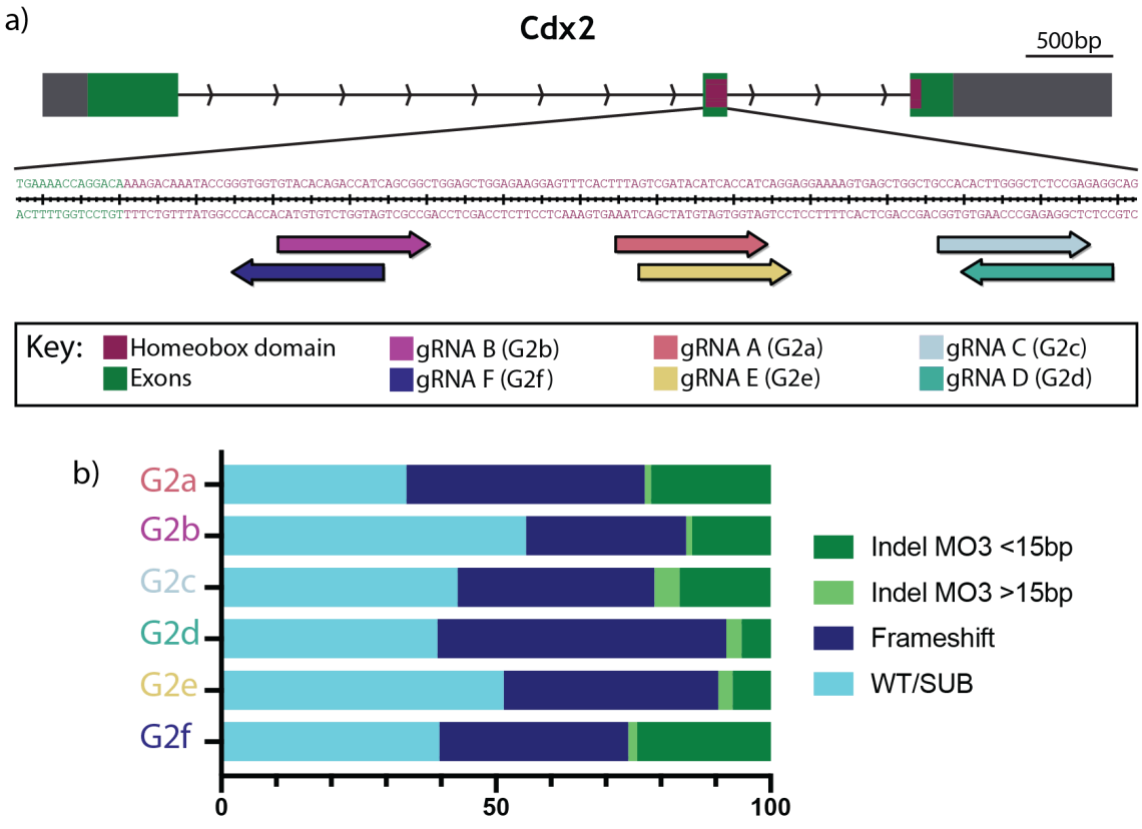



**Figure 6.1: Mutation efficiency of single gRNAs targeting the *Cdx2* homeobox is insufficient.** a) Schematic outlining the position of each gRNA designed to target the *Cdx2* homeobox domain in exon 2. b) Bar chart showing the mutagenic efficiency of each single gRNAs. Transfected cells were isolated by FACS, the target region amplified from bulk gDNA and sequenced using MiSeq. Using this data, it is possible to determine any allele variants generated by CRISPR targeting and their frequency within the population. Alleles were classified as the following genotypes: wild-type or substitution (WT/SUB), frameshift, indels that were multiples of three (MO3) that were less than 15bp in length and MO3 indels that were larger than 15bp.

Most of the *Cdx2* homeobox domain is located in exon 2 (Figure 6.1a). Six gRNAs (G2a-f) targeting this region were identified using the MIT CRISPR Designer and were predicted to have low off-target effects (Figure 6.1a) (Cong et al., 2013). Each guide was cloned into a plasmid (px330) containing a human-coding optimised spCas9, a trans-activating crRNA element and an mCherry cassette. To determine their efficiency, TSCs were transfected with a gRNA-containing plasmid or with an empty plasmid as a control. I aimed to mutate >75% of alleles in each polyclonal population as this would imply that over 50% of cells were homozygous for a mutation within a diploid population.

Transfected cells were isolated based on their mCherry expression 24-hours post-transfection using FACS as previously described (Chapter 4.1, Figure 4.1) and they were genotyped by MiSeq sequencing. Cas-Analyzer was used both to classify the different allele variants based on the type of mutation and to determine their frequencies (Park et al., 2017). This tool categorises point mutations as 'substitute' and groups them with wild-type alleles (WT/SUB). As the effect of a point mutation would be difficult to predict, it was decided that only deletion and insertions would be considered as mutations during analyses. As expected, INX controls samples contained 99.9% WT alleles (data not shown). No gRNA induced mutations in >75% of alleles as desired, but G2a, G2c, G2d and G2f were the most mutagenic (Figure 6.1b).

To increase the mutation efficiency, cells were co-transfected with two gRNA vector combinations; G2a + G2c (A+C) and G2a + G2c (A+D) (Figure 6.2a). Both combinations reproducibly increased mutagenic efficiency, although the A+D combination was less efficient in Rossant-GFP TSCs (27.5% WT/SUB) than INX TSCs (15% WT/SUB) (Figure 6.2a-b). These gRNA combinations also reproducibly induced the same type of mutations, generating two common mutations that accounted for 44.2-59.1% of total A+C alleles (Figure 6.2c) and 57.2-71.1% of total A+D alleles (Figure 6.2d) across cell lines and biological replicates. Therefore, both gRNA combinations were sufficiently mutagenic to proceed with validating the effect of *Cdx2* mutations on polyclonal populations.

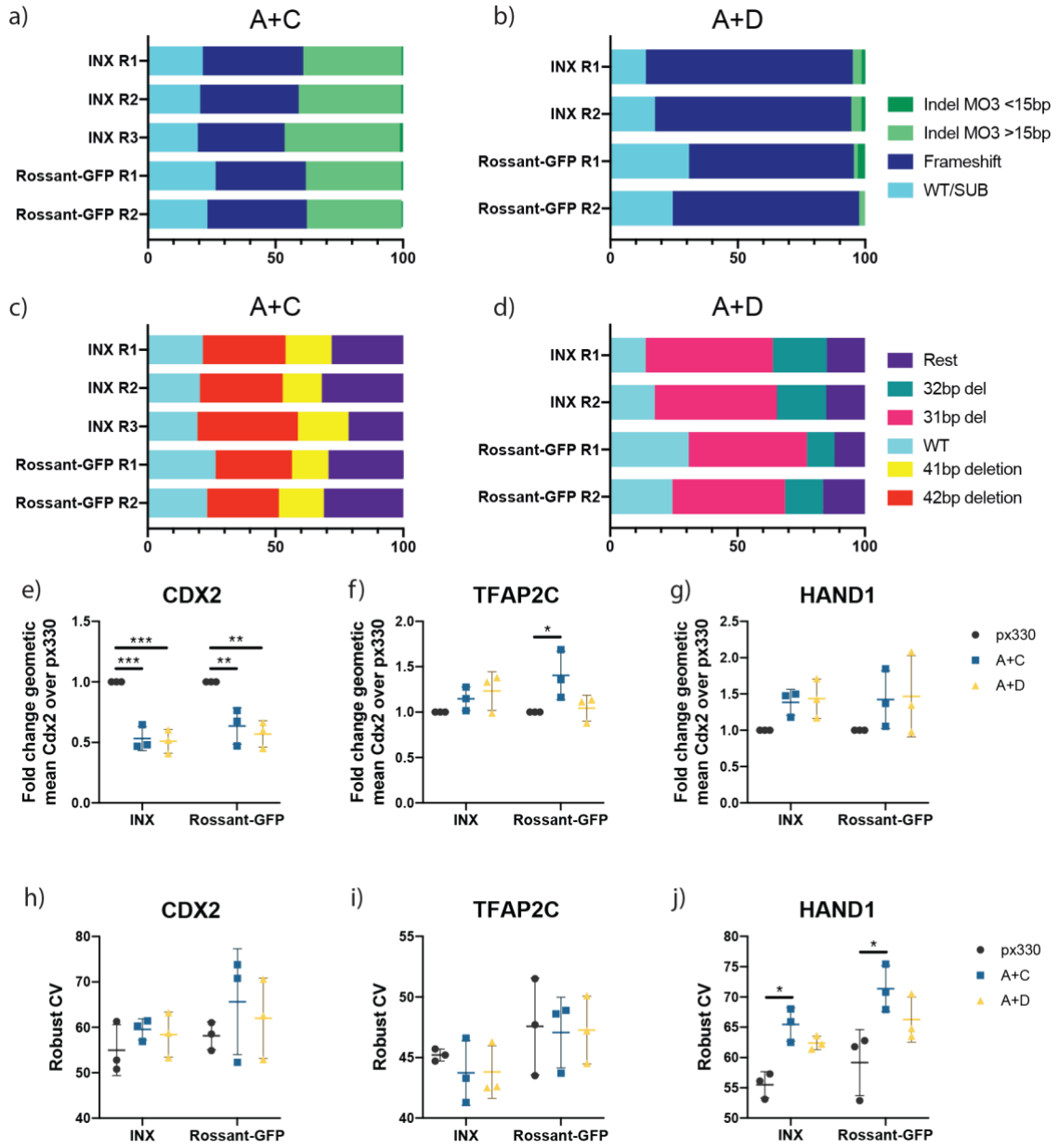



**Figure 6.2: A+C and A+D gRNA combinations generate reproducible mutations and are sufficiently mutagenic.** a-d) TSCs were transfected with gRNA pairs G2a and G2c (A+C) and G2a and G2d (A+D). Positively transfected cells were isolated by FACS 24-hours post-transfection and genotyped by MiSeq. Each bar represents one biological replicate. a-b) Bar charts displaying allele variants generated by A+C (a) and A+D (b) classified into the following genotypes: wild-type or substitute (WT/SUB), frameshift, indels that were multiples of three (MO3) less than 15bp in length and MO3 indels that were larger than 15bp. c-d) Bar charts to represent proportion of reads attributed to the three most common alleles in A+C (c) and A+D (d) polyclonal populations. e-f) Fold change in geometric mean expression of CDX2 (e), TFAP2C (f) and HAND1 (g) expression in A+C and A+D over px330. Analysis performed by flow cytometry (n=3 biological replicates per cell line). Error bars display SD. Statistical significance was determined by two-way ANOVA tests: * $p \leq 0.0332$, ** $p \leq 0.0021$, *** $p \leq 0.0002$. h-i) Robust coefficient of variation (CV) of CDX2 (h), TFAP2C (i) and HAND1 (j) expression in A+C, A+D and px330 determined by FlowJo. n=3 biological replicates per cell line. Error bars display SD. Two-way ANOVA test: * $p \leq 0.0332$.

Next, I aimed to determine how efficiently *Cdx2* mutations reduce CDX2 expression 48-hours post-transfection (Figure 6.2e). CDX2 expression was reduced by approximately 49% in all INX mutant experiments (Figure 6.2e); a level similar to that observed in knockdown experiments (55%, Chapter 4.1, Figure 4.1c). In line with the increased presence of WT/SUB alleles in these samples, CDX2 expression was less reduced in Rossant-GFP TSCs than INX TSCs for both the A+C (36.4%) and A+D (43%) combinations (Figure 6.2e). How efficiently *Cdx2* mutant populations reduced CDX2 expression changed between biological replicates for both Rossant-GFP and INX TSCs, perhaps because some cells in these populations retain WT alleles. To assess variability in the expression of CDX2 in cells of the same sample, I measured the coefficient of variation (CV): the dispersion of the data around the mean (Figure 6.2h). Although the variation was similar between different INX conditions, Rossant-GFP A+C and A+D samples were generally more variable than px330 control samples. In all, this suggests that the reduction in CDX2 expression in *Cdx2* mutant polyclonal cells can be heterogeneous and is probably related to the efficiency of inducing *Cdx2* mutations.

In Chapters 4 and 5, I showed that HAND1 and TFAP2C upregulation are key drivers of aberrant TSCs differentiation in long-term *Cdx2* knockdown. Whilst *Cdx2* mutant polyclonal lines showed a trend in upregulating TFAP2C and HAND1 expression 48-hours post-transfection, this was not statistically significant in most experiments (Figure 6.2f-g). The observed variability in both HAND1 and TFAP2C expression may be related to the differences in *Cdx2* mutation efficiency between biological replicates. Therefore, I measured the robust CV of both proteins (Figure 6.2f-g). HAND1 expression is particularly variable between cells in *Cdx2* mutant conditions, suggesting that some cells do upregulated its expression, whilst others do not. Overall, HAND1 and TFAP2C expression are upregulated when *Cdx2* is mutated, but their expression is variable between cells in polyclonal population.

## 6.2 Mutated alleles are outcompeted by wild-type alleles in bulk Cdx2 mutant populations

To assess the long-term effect of *Cdx2* mutations on TSCs, transfected A+C, A+D and px330 cells were isolated using FACS 24-hours after transfection to generate polyclonal cell lines. By 144-hours, there was a small increase in the number of TGCs in A+C and A+D cultures (data not shown), but this did not recapitulate the homogeneous phenotype observed after long-term *Cdx2* knockdown (Chapter 5.1). At 24-hours post-transfection, up to 25% of alleles in these polyclonal populations were WT/SUB alleles (Figure 6.2a-b). To determine whether this lack of long-term phenotype reflected a selective advantage for cells containing wild-type alleles, the frequency of mutations in polyclonal populations was determined at 96-hours and 144-hours post-transfection (Figure 6.3a-b).

In A+C polyclonal populations, the proportion of WT/SUB alleles increased between each successive time point in both INX and Rossant-GFP polyclonal populations (Figure 6.3a). In contrast, the proportion of WT/SUB alleles did not change between 24-hours and 96-hours after transfection in A+D Rossant-GFP (27.6% versus 27.6%) and A+D INX (15.6% versus 18.1%) polyclonal populations (Figure 6.3b). However, by 144-hours after transfection, 47.7% and 44.2% of total alleles in A+D Rossant-GFP and INX polyclonal cell lines were WT/SUB, respectively (Figure 6.3b).

To discover whether specific mutations were being outcompeted in the polyclonal populations, I assessed the percentage of reads attributed to the three most common alleles over time (Figure 6.3c-d). In A+C polyclonal cultures, the 42bp and 41bp deletion alleles were lost in similar proportions to the ‘rest’ of the edited alleles in both cell lines over time (Figure 6.3c). In the A+D polyclonal lines, there was a 32.7% and 38.8% reduction in the proportion of the most dominant allele, the 31bp deletion, between 24-hours and 144-hours post-transfection for Rossant-GFP and INX TSCs, respectively (Figure 6.3d). Perhaps because of their lower overall contribution, changes to the proportion of the 32bp deletion and ‘rest’ alleles in A+D populations were smaller and less consistent (Figure 6.3d). In all, it appears that *Cdx2* mutant alleles are outcompeted from polyclonal populations by WT alleles at similar rates.

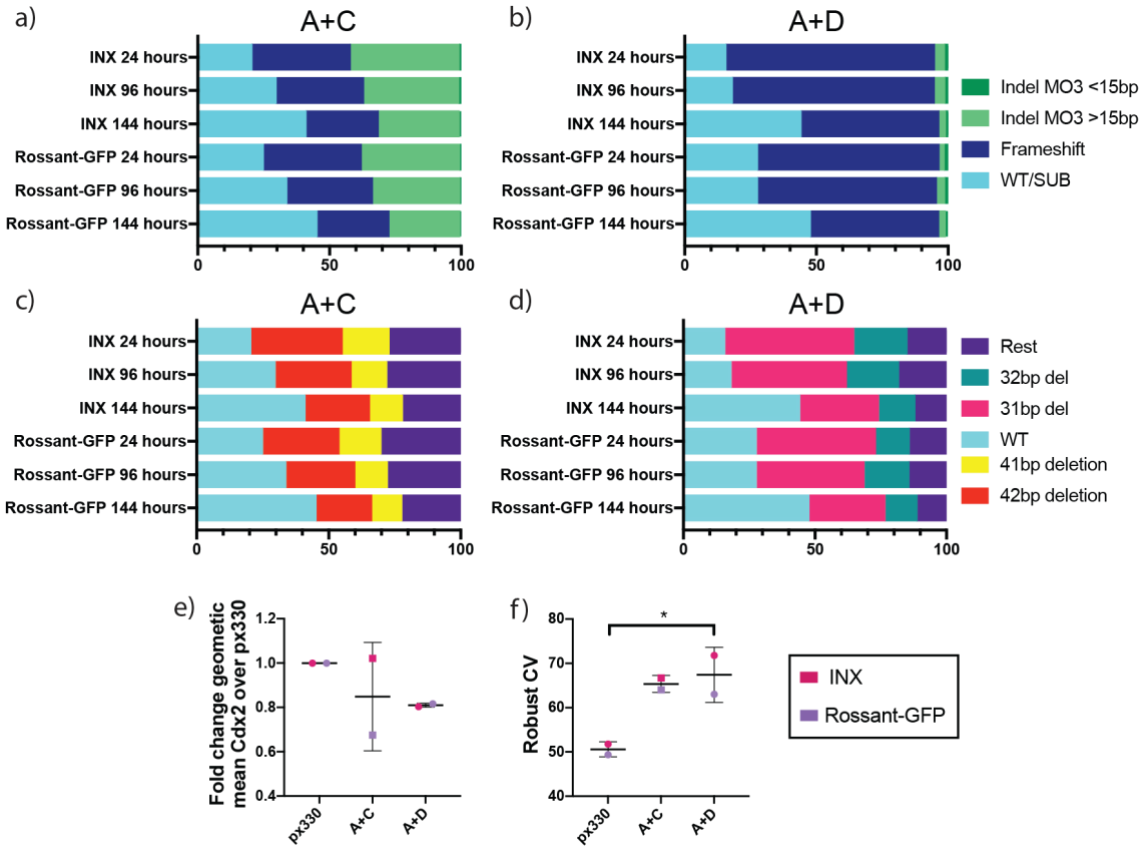



**Figure 6.3: Wild-type alleles outcompete mutant alleles in *Cdx2*-targeted polyclonal populations.** a-d) Positively transfected TSCs with gRNA pairs A+C or A+D were isolated by FACS 24-hours post-transfection. Cells were either used for genotyping by MiSeq immediately (24 hours) or replated in bulk and collected for genotyping at 96-hour or 144-hours post-transfection. n=2 biological replicates for 24-hour genotypes; n=1 biological replicate for 96-hour and 144-hour genotypes. a-b) Bar charts displaying allele variants frequencies A+C (a) and A+D (b) polyclonal populations classified into the following genotypes: wild-type or substitute (WT/SUB), frameshift, indels less than 15bp in length that were multiples of 3 (MO3) and MO3 indels that were larger than 15bp. c-d) Bar charts represent proportion of reads attributed to the three most common alleles found within A+C (c) and A+D (d) populations. e) Fold change in geometric mean expression of CDX2 in *Cdx2* mutant A+C and A+D polyclonal populations (n=1 biological replicate per cell line) over the relevant cell line px330. Analysis was performed by flow cytometry. Error bars display SD. f) Robust CV of CDX2 expression in *Cdx2* mutant A+C and A+D and px330 polyclonal populations determined by FlowJo. (n=1 biological replicate A+C, A+D and px330 per cell line.) Error bars display SD. Significance determined by one-way ANOVA: * $p \leq 0.0332$.

Given the increase in WT alleles in all polyclonal edited populations over time, I next examined the expression of CDX2 in these cells at 144-hours post-transfection. Although CDX2 expression remained up to 32% lower in mutant clones than in px330 controls (Figure 6.3e), this still represented an overall increase in CDX2 expression compared to its expression 48-hours post-transfection, where it was 36-55% lower than px330 controls (Figure 6.2e). Further, there was significant variation in CDX2 expression between cells in *Cdx2* targeted populations at 144-hours (Figure 6.3f). In all, this confirms that there is a significant increase in the number of cells in polyclonal populations expressing higher levels of CDX2 by 144-hours post-transfection.

## 6.3 Clonal *Cdx2* mutant TSCs can be maintained for several passages but are inherently unstable

*Cdx2*-mutant alleles are outcompeted by WT alleles in polyclonal populations. To determine the effect of *Cdx2* mutations on TSCs, clonal analysis was performed.

### 6.3.1 Generation, characterisation and genotyping of single-cell CRISPR clones

To generate clonal lines, INX and Rossant-GFP TSCs were transfected with either the optimised gRNA combinations (A+C, A+D) or an empty vector control (px330). Positively transfected cells were single-cell sorted into 96 well plates 24-hours after transfection. Clones were categorised based on their appearance seven days after being single cell sorted as either wild-type ‘colony’, ‘colony differentiating’ or ‘TGCs’ (Figure 6.4a).

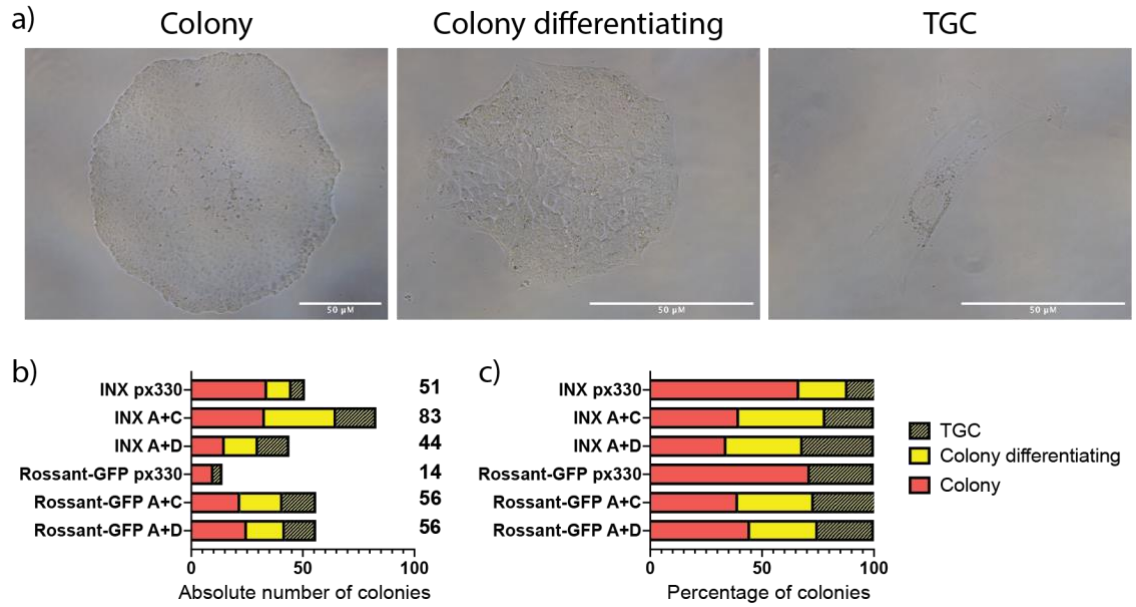



**Figure 6.4: *Cdx2*-targeted TSCs can form clonal trophoblast outgrowths.** a) Representative brightfield images taken of clonal outgrowths classified as ‘Colony’, ‘Colony differentiating’ and ‘Trophoblast giant cell (TGC)’ 7 days post-single cell sort. Colonies that showed morphological characteristics of TSCs with low levels of visible differentiation were categorised as ‘colony’. Colonies that showed substantial differentiation, including large colonies of TGCs, were categorised as ‘colony differentiating’ so as to allow for the possibility of surviving stem cells. When wells only contained a few TGCs, they were categorised as ‘TGCs’. b) Bar chart to display the total number of outgrowths and how they were classified as described in (a). c) To enable classification of colonies to be compared between conditions and cell lines, the number of clones in each category is displayed as a proportion of the total for each cell line and condition.

In total, I picked between 14 and 83 clones per condition and cell line across two experiments (Figure 6.4b). Increased exogenous stress, such as transfection and/or cell sorting, can result in spontaneous TSC differentiation. Thus, it was unsurprising to see 'differentiating colony' or 'TGCs' clones in control px330 clonal experiments (~30%). Despite this, the proportion of clones categorised as 'TGC' or 'colony differentiating' was higher in all *Cdx2*-targeting conditions (55.4-65.9%) relative to the px330 control regardless of the cell line used (Figure 6.4c).

Attempts to expand clones over multiple passages to generate clonal cell lines were not successful for all clones (Figure 6.5a-d). Only one TGC clone (striped yellow bars, Figure 6.5a-d) and a large proportion of 'colony differentiating' clones (50-94%, yellow bars, Figure 6.5a-d) survived to passage three. This initial segregation of clones into 'TGC' or 'colony differentiating' categories was, therefore, predictive of their reduced survival rate. In contrast, the majority of wild-type colony clones survived to passage three (72-90.9%, Figure 6.5a-d).

It was noted that some wild-type clones changed in morphology over several passages, indicating that they were differentiating. These clones may be edited clones whose phenotype is less pronounced in early passages. Therefore, the clones were re-classified based on their appearance at passage three (Figure 6.5e). This analysis revealed that a large number (72-86.7%) of clones derived from gRNA edited populations show some level of differentiation by passage three (Figure 6.5e).

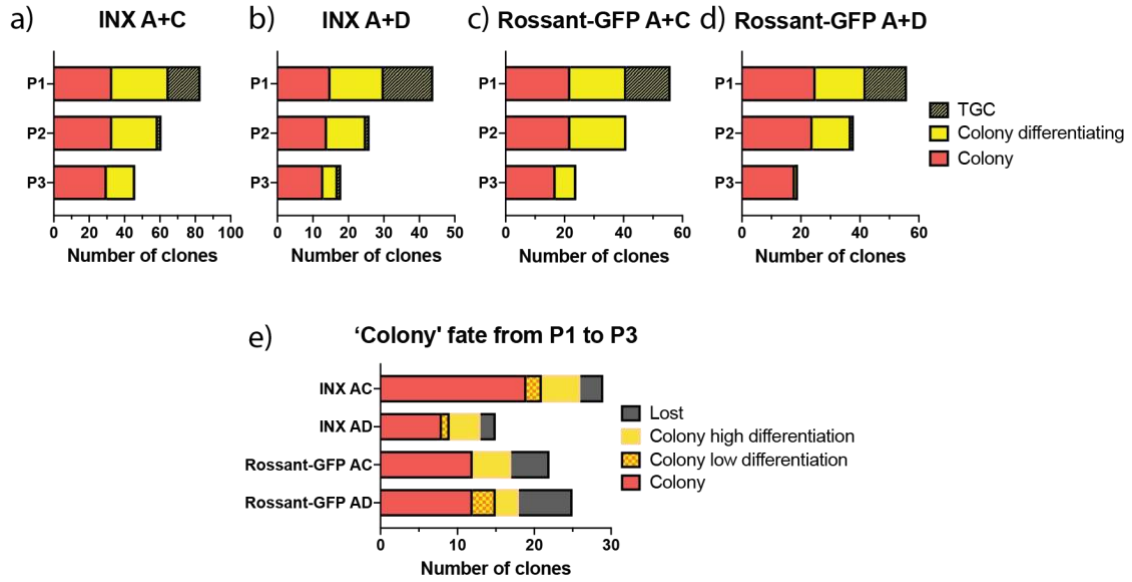



**Figure 6.5: 'Colony' and some 'colony differentiating' *Cdx2*-targeted clonal populations can be expanded over multiple passages.** a-d) Attempts were made to passage all clonal outgrowths, but many clones were lost over successive passages. To assess which classification of clones were most frequently lost, the number of clones that contained any cells at passage two (P2) and passage three (P3) were determined. These bar charts display the total number of clones that remained in culture for each of the original categories from passage 1 (P1) to P3. e) The majority of clones classified as 'colony' at P1 were successfully passaged to P3, but their appearance changed. This bar chart displays the fate of all clones categorised as 'colony' at P1 and their re-classification at P3.

I next focused on genotyping these characterised clones. Sanger sequencing revealed that many of these clones contained compound mutations and some were either polyploid or polyclonal because they possessed three alleles. Whilst TSC lines are diploid in early passages, most TSC lines tested display some form of aneuploidy after passage 40 regardless of the culture conditions (Kubaczka et al., 2014). All experiments in this thesis used INX TSCs at passage 55 and above and the passage number of the Rossant-GFP TSC line was unknown. Whole exome sequencing (WES) for these TSCs lines confirmed that they both are aneuploid (data not shown, personal correspondence with Daniel Snell).

To precisely characterise the mutations and their ratios in each clone, genotyping was performed using a multiplexed MiSeq approach with three primer sets (see 2.2.10.1 and Figure 6.6a). Clones were grouped into categories according to the type of mutation: wild-type, frameshift or large deletions. When a clone contained two different alleles in equal proportions, it was classified as compound heterozygous. If the difference between the frequencies of two alleles was greater than 10%, or if a clone contained more than two alleles, it was classified as >2N.

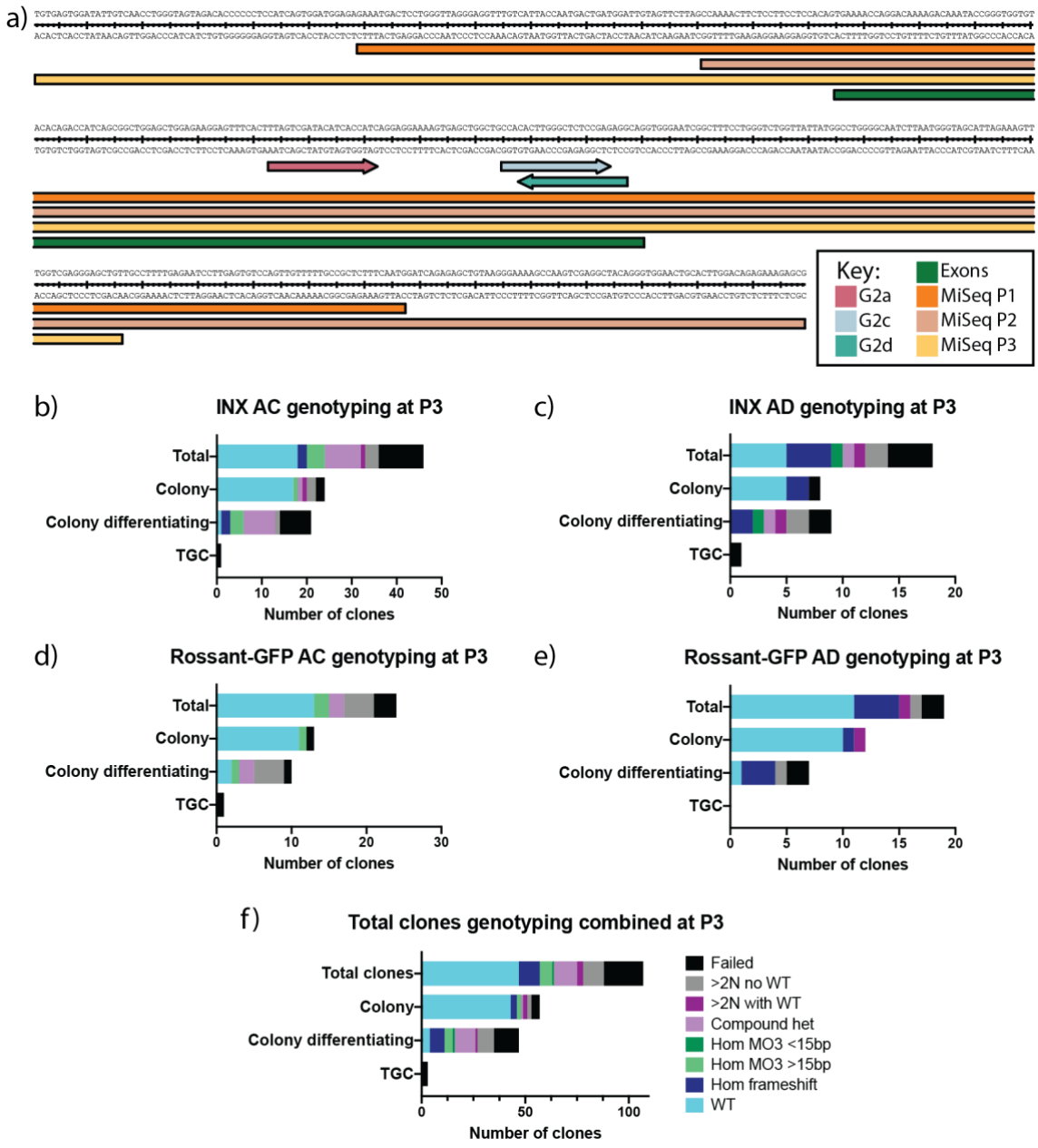



**Figure 6.6: Genotyping of *Cdx2*-targeted clones reveals that they can be expanded in the absence of wild-type *Cdx2* alleles.** a) A schematic outlining the multiplex MiSeq approach used to genotype clonal lines. gDNA from each clone was PCR amplified using one of the three MiSeq primer sets outlined. PCR amplicons from one of each primer set were combined for MiSeq library preparation and sequencing. b-f) MiSeq was used to genotype clones. Genotypes are provided using overall categories: wild-type (WT), Homozygous (hom), multiple of 3 (MO3), heterozygous (het). Compound hets were clones with two alleles in approximate 50:50 proportions. Over 2N (>2N) represents clones where difference between the frequencies of two alleles was greater than 10% or where there were more than two alleles detected. Failed clones were those whose sequencing was poor or the reads number was insufficient due to low gDNA amounts extracted from clones. Genotypes are provided with information as to the phenotypic appearance of clones at P3 according to the criteria outlined in Figure 6.4a.

In line with the poorer mutation efficacy observed in Rossant-GFP TSCs compared with INX TSCs (Figure 6.3), a higher proportion of Rossant-GFP clones were wild-type when compared to INX clones (Figure 6.6b-e). Across both cell lines and gRNA pairs, 75.4% of clones that were classified as morphologically normal TSCs 'colonies' at passage three (Figure 6.5e) were genotyped as wild-type and 79% contained least one wild-type allele (Figure 6.6f). In contrast, only 10.6% of clones that were classified as 'colony differentiating' contained wild-type alleles (Figure 6.6f). Therefore, the morphological appearance of clones could be used to accurately predict between wild-type and edited clones. There was insufficient genomic material to genotype 25.5% of the 'differentiating colony' clones and all the 'TGC' terminally differentiated clones (black bars, Figure 6.6b-f). These 'failed' clones were lost during subsequent passages due to differentiation (data not shown), suggesting that *Cdx2* mutations destabilised the cells and resulted in terminal differentiation.

### 6.3.2 Not all clones are born equal: the same mutation does not affect TSCs the same way

Unlike *Cdx2* knockdown cells, clones containing only *Cdx2* mutant alleles can be maintained, although they are prone to differentiate. The rate at which *Cdx2* mutant clones differentiate and/or are lost from the population is different. One potential explanation for these phenotypic differences is that different mutations compromise CDX2 function differently. With this in mind, I next asked if specific phenotypes were associated with unique genotypes (Table 6.1).

| Number | Genotype | Colony | Colony Differentiating | Predicted effect |
|---|---|---|---|---|
| 1 | 103bp del + 78bp del (3:1) | | 1 | Frameshift and in-frame |
| 2 | 110bp del homozygous | | 1 | Frameshift |
| 3 | 31bp del + 1bp ins (3:2) | | 1 | Frameshift |
| 4 | 31bp del homozgous | 2 | 5 | Frameshift |
| 5 | 32bp del + 31bp del (1:3) / (3:1) | | 3 | Frameshift |
| 6 | 32bp del homozygous | | 1 | Frameshift |
| 7 | 33bp del homozygous | | 1 | In-frame |
| 8 | 35bp del + 3bp del (3:2) | 1 | | In-frame |
| 9 | 38bp del + 35bp del (3:1) | | 1 | Frameshift and in-frame |
| 10 | 3bp del homozygous | | 1 | In-frame |
| 11 | 41bp del + 37bp del (1:1) | | 1 | Frameshift |
| 12 | 41bp del homozygous | | 1 | Frameshift |
| 13 | 42bp del + 33bp del (1:1) | | 1 | In-frame |
| 14 | 42bp del + 35bp del (1:1) | | 1 | In-frame |
| 15 | 42bp del + 37bp del (1:1) | | 2 | Frameshift and in-frame |
| 16 | 42bp del + 37bp del + 41bp del (1:1:1) | 1 | | Frameshift and in-frame |
| 17 | 42bp del + 38bp del (1:1) / (3:1) | 1 | 1 | Frameshift and in-frame |
| 18 | 42bp del + 39bp del (1:1) | | 1 | In-frame |
| 19 | 42bp del + 40bp del (1:1) | | 1 | Frameshift and in-frame |
| 20 | 42bp del + 41bp del (1:1) / (1:2) / (3:1) | | 3 | Frameshift and in-frame |
| 21 | 42bp del + 45bp del (1:1) | | 1 | In-frame |
| 22 | 42bp del homozygous | 2 | 3 | In-frame |
| 23 | WT + 1bp ins (4:1) | 1 | | WT and frameshift |
| 24 | WT + 32bp del + 6bp del (3:3:2) | | 1 | WT and inframe |
| 25 | WT + 35bp del (2:1) | 1 | | WT and inframe |

**Table 6.1: Most *Cdx2*-mutant clones that show no visible phenotype at passage three contain an in-frame mutation.** This table summarises all of the unique genotypes found for clones in both A+C and A+D gRNA combinations in the INX and Rossant-GFP cell lines. The phenotypic appearance of each clone at passage 3 was classified as colony or colony differentiating as described in Figure 6.4a. All genotypes are categorised as in-frame or frameshift.

CDX2 contains a highly conserved 60 amino acid homeobox domain (Kissinger et al., 1990, Dragan et al., 2006, Burglin and Affolter, 2016). These 60 amino acids encode three helices and it is the third and longest helix that interacts with DNA (Burglin and Affolter, 2016). Recent work has shown that the human CDX2 protein preferentially binds methylated DNA using suitably positioned hydrophobic residues, Ile232 and Ala239, within this third helix (Yin et al., 2017).

The two most common alleles in surviving clones were a 42bp deletion and a 31bp deletion, generated through editing with A+C or A+D gRNA combinations, respectively (Table 6.1). Both deletions are predicted to start 81bp into the homeobox domain, at amino acid 27. The 42bp deletion is predicted to cause an in-frame mutation that would result in deletion of the subsequent 14 amino acids. This 14 amino acid region contains all of helix two and the loop region to helix 3 (Kissinger et al., 1990, Burglin and Affolter, 2016). It is unlikely, although not impossible, that CDX2 could function without such a large and important structural compartment of

this highly conserved domain. The 31bp deletion is predicted to cause a frameshift that would result in a premature stop codon in exon 3 and reduce the protein size by 42 amino acids. This, as with all frame-shift mutations listed in Table 6.1, would alter the amino acid sequence of the third helix of the homeodomain and affect the DNA binding function of CDX2. As such, clones with large deletions or with frameshift mutations were predicted to have a complete loss of CDX2 function. In addition to these mutations, only a few clones had small in-frame deletions of only a few amino acids in length (8, 10 and 24, Table 6.1). Whether these are sufficient to disrupt CDX2 function is unclear.

In all, a diverse range of phenotypes were observed in *Cdx2* mutant clones irrespective of their genotype (Table 6.1). There was no visible differentiation in both clones where over 60% of alleles are wild-type (23 and 24, Table 6.1), suggesting that *Cdx2* expression over a certain threshold maintains undifferentiated TSCs. However, many clones had no wild-type alleles and showed no obvious increased differentiation. Interestingly, 71.4% of these morphologically normal *Cdx2* mutant clones (8, 16, 17 and 22, Table 6.1) contained at least one copy of an in-frame mutation, either 42bp, 35bp or 3bp deletion, suggesting that in-frame mutations might generate a functional CDX2 protein. Beyond this, despite containing identical homozygous mutations, 28.6% of 31bp deletion clones (4, Table 6.1) and 40% of 42bp deletion clones (22, Table 6.1) showed no obvious increase in differentiation at passage three. Therefore, the same mutation in the same cell lines does not necessarily induce the same phenotype.

### 6.3.3 Spot the difference: linking genotype to phenotype

The different phenotypes of *Cdx2* mutant clones with the same homozygous mutations are intriguing. I next validated the genotype of multiple wild-type and edited (compound heterozygous (CH), frameshift (FS) and large deletion (LD)) clones by sequencing their full-length *Cdx2* cDNA (Table 6.2). Several RNA samples had low yields of RNA and were excluded from analysis. Sanger sequencing of these cDNAs revealed that some clones contained more than one *Cdx2* transcript. Where more than one transcript was detected, PCR products were TA cloned and re-sequenced.

| Experiment no. | Clone name | Genotype by MiSeq | Genotype confirmed by cDNA | Alternative splicing? | Passage no. at cryopreservation | Phenotype at cryopreservation |
|---|---|---|---|---|---|---|
| 1 | INX-CH-1 | 42bp del + 37bp del | Yes | No | P6 | Mostly TGCs + small WT colonies |
| 1 | INX-CH-2 | 42bp del + 40bp del | Yes | Presence of exon skipping (146bp del) | P6 | TGC only |
| 2 | INX-CH-3 | 42bp del + 38bp del | Not confirmed<br>Sequencing failed | Presence of exon skipping (146bp del)<br>Alternative splicing variant using intronic sequence | P3 | WT colonies |
| 2 | INX-CH-4 | 42bp del + 41bp del | No 41bp del detected | None detected. Sequencing failed | P3 | Differentiating + dying |
| 2 | INX-CH-5 | 42bp del + 39bp del | Yes | Presence of exon skipping (146bp del) | P3 | WT colonies + small numbers of TGCs |
| 2 | INX-CH-6 | 42bp del + 41bp del + 37bp del | No 42bp del detected | Presence of exon skipping (146bp del) | P3 | TGCs + colonies |
| 1 | INX-FS-1 | 31bp del hom | Yes | No | P7 | Mostly TGCs + small WT colonies |
| 1 | INX-FS-2 | 32bp del hom | Yes | No | P6 | Mostly TGCs + small WT colonies |
| 1 | INX-FS-3 | 41bp del hom | Yes | Presence of exon skipping (146bp del) | P6 | Mostly TGCs + small WT colonies |
| 2 | INX-FS-4 | 31bp del hom | Yes | No | P3 | WT colonies + small numbers of TGCs |
| 2 | INX-FS-5 | 31bp del hom | Yes | No | P3 | WT colonies + small numbers of TGCs |
| 1 | INX-LD-1 | 42bp del hom | Yes | Presence of exon skipping (146bp del)<br>Alternative splicing variant using intronic sequence | P6 | WT colonies + TGCs |
| 1 | INX-LD-2 | 42bp del hom | Yes | Presence of exon skipping (146bp del)<br>Alternative splicing variant using another sequence | P6 | Mostly TGCs + small WT colonies |
| 2 | INX-LD-3 | 33bp del hom | Yes | Presence of exon skipping (146bp del) | P3 | WT colonies + TGCs |
| 2 | INX-LD-4 | 42bp del hom | Yes | No | P3 | WT colonies + TGCs |
| 2 | INX-LD-5 | 42bp del hom | Yes | No | P3 | TGC only |
| 1 | INX-WT-1 | WT | Yes | No | P7 | WT colonies |
| 1 | INX-WT-2 | WT | Yes | No | P7 | WT colonies |
| 1 | INX-WT-3 | WT | Yes | No | P7 | WT colonies |
| 2 | INX-WT-4 | WT | Yes | No | P3 | WT colonies |
| 2 | INX-WT-5 | WT | Yes | No | P3 | WT colonies |
| 2 | ROSS-CH-1 | 42bp del + 41bp del | Yes | No | P3 | Mostly TGC + small WT colonies |
| 2 | ROSS-CH-2 | 42bp del + 35bpp del | Yes | No | P3 | WT colonies + TGCs |
| 2 | ROSS-FS-1 | 31bp del hom | Yes | No | P3 | WT colonies + TGCs |
| 2 | ROSS-FS-2 | 31bp del hom | Yes | No | P3 | Mostly TGC + small WT colonies |
| 1 | ROSS-LD-1 | 42bp del hom | Yes | No | P6 | Mostly TGC + small WT colonies |
| 2 | ROSS-LD-2 | 42bp del hom | Yes | No | P3 | WT colonies + TGCs |
| 1 | ROSS-WT-1 | WT | Yes | No | P7 | WT colonies + small numbers of TGCs |
| 1 | ROSS-WT-2 | WT | Yes | No | P7 | WT colonies |
| 1 | ROSS-WT-3 | WT | Yes | No | P7 | WT colonies |
| 2 | ROSS-WT-4 | WT | Yes | No | P3 | WT colonies |
| 2 | ROSS-WT-5 | WT | Yes | No | P3 | WT colonies |

**Table 6.2: Genotype, transcript validation and morphological appearance of *Cdx2* mutant clones.**

Some INX clones contained alternatively spliced *Cdx2* transcripts (Table 6.2). In total, 50% (8/16) of all INX mutant clones analysed contained an alternative spliced transcript in which exon 2 was skipped. To model the loss of *Cdx2 in vivo*, mouse models have been generated whereby exon 2 is floxed and can be removed by Cre recombination. The resulting transcript is the same as observed here (van de Ven et al., 2011), indicating that this generates a non-functional CDX2 protein. Other alternative splicing *Cdx2* variants were also found in clones INX-CH-3, INX-LD-1 and INX-LD-2. No alternative splicing was observed for INX clones containing a 31bp or a 32bp deletion, suggesting that only mutations larger than these affect splicing.

Interestingly, no mutant Rossant-GFP clones showed any alternative splicing, even for deletions larger than 32bp. Overall, Rossant-GFP *Cdx2* mutant clone survival was poorer than INX clones, with only one validated Rossant-GFP mutant clone reaching passage six (Table 6.2). The lack of alternative splicing in Rossant-GFP mutant clones may represent a biological difference in the ability of INX and Rossant-GFP cell lines to compensate for *Cdx2* loss.

Previous work has shown that CRISPR-Cas9 genome editing can cause exon skipping, probably as a result of mutations in splice-regulatory sequences (Kapahnke et al., 2016, Mou et al., 2017). Exonic splicing enhancers (ESE), which represent binding sites for highly conserved splicing factors (serine/arginine (SR) proteins), are common in exons and are important for regulating alternative splicing (Cartegni et al., 2003). At these sites, SR proteins recruit the splicing machinery and/or antagonise repressive elements in the vicinity (Cartegni et al., 2002, Graveley, 2000).

Therefore, I next examined the distribution of putative ESE motifs across *Cdx2* exon 2 using the online tool ‘ESE finder 3.0’ (Figure 6.7) (Cartegni et al., 2003, Smith et al., 2006). As expected, ESE motifs are common across exon 2, many of which would be unaffected in *Cdx2* mutant clones. At least one motif for every SR protein assessed, except SRSF6, was present in the region known to be affected by the two most common mutations generated in *Cdx2* mutant clones: the 31bp deletion and the 42bp deletion. Furthermore, there are specific motifs that would be uniquely affected in mutations larger than the 31bp deletion (Figure 6.7). Therefore, the lack

of one or more of these putative motifs at the end of exon 2 may be responsible for the alternative splicing seen in edited clones with mutations larger than 31bp.

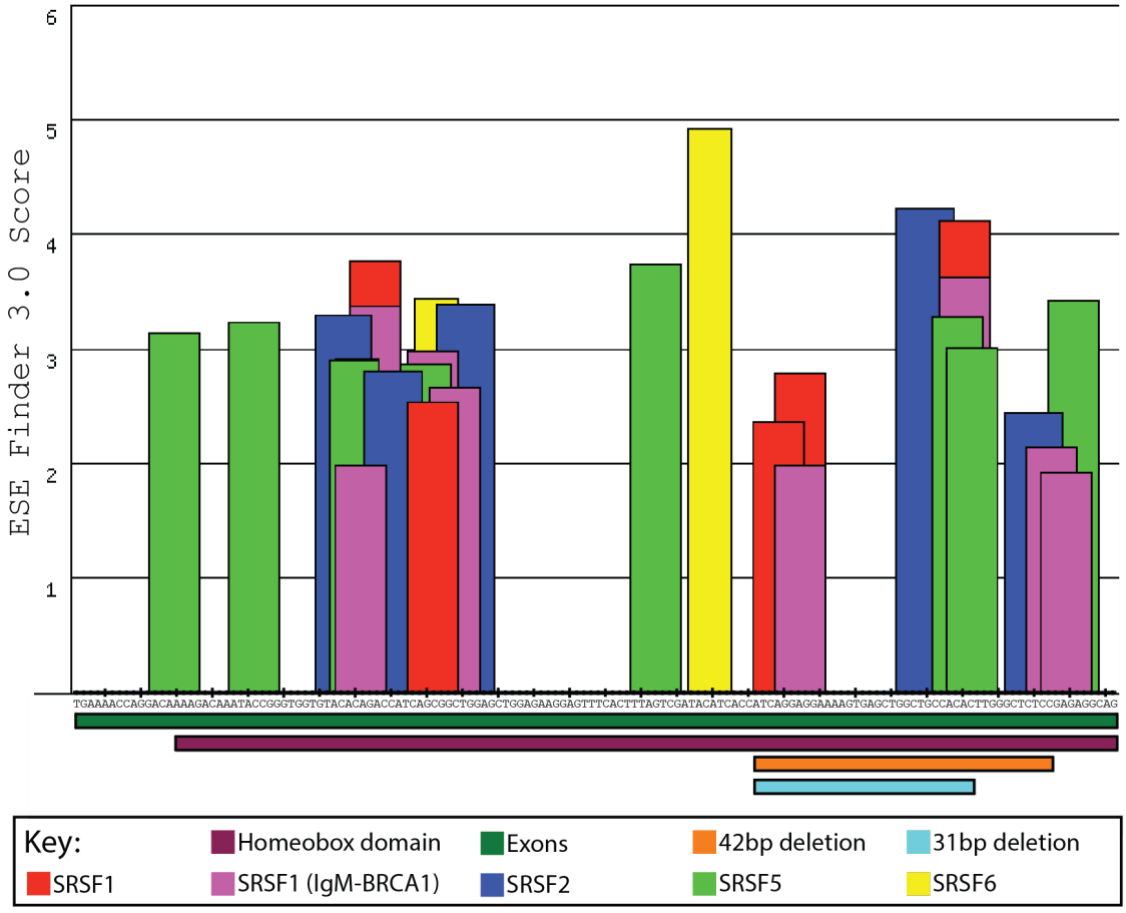



**Figure 6.7: Positioning of exonic splicing enhancers (ESEs) across *Cdx2* exon 2.** The positions of conserved motifs for SR proteins within exon 2 of *Cdx2* as determined by 'ESE finder 3.0'. The distribution of these motifs is shown relative to the positions of the highly prevalent 31bp and 42bp deletions.

### 6.3.4 Characterising the gene regulatory network of *Cdx2*-mutant clones

In contrast to *Cdx2* knockdown cells, some *Cdx2* mutant clones can be maintained for a few passages. However, cell morphology indicates that some clones are unstable and contain terminally differentiated cells. One potential explanation for the longer survival of mutant clones, especially those with frameshift and larger deletions, is that these clones use transcriptional adaptations. Upon knockout of a gene, it is known that homologous genes can become misregulated in an apparent attempt to compensate for the loss of the mutated gene (reviewed in Kafri et al., 2009). However, as these characterised mutations either do not introduce frameshifts or do not induce a premature termination codon until the last exon, it is

unlikely that *Cdx2* transcripts would undergo nonsense mediated decay (NMD) (Lindeboom et al., 2016). NMD is thought to be required for transcriptional adaptation (El-Brolosy et al., 2019).

Nonetheless, *Cdx2* has two homologs: *Cdx1* and *Cdx4*. Although neither homolog is expressed in the trophoblast lineage, work completed by Niwa et al (2005) demonstrated that overexpression of *Cdx4*, but not *Cdx1*, can transdifferentiate mESCs to TSC-like cells (Niwa et al., 2005). Given this, it is possible that *Cdx2* mutant clones may survive because of genetic compensation. To test this hypothesis, I examined the expression of *Cdx2*, using primers upstream or downstream of exon 2 in order to pick up alternative spliced variants, as well as *Cdx1* and *Cdx4* (Figure 6.8).

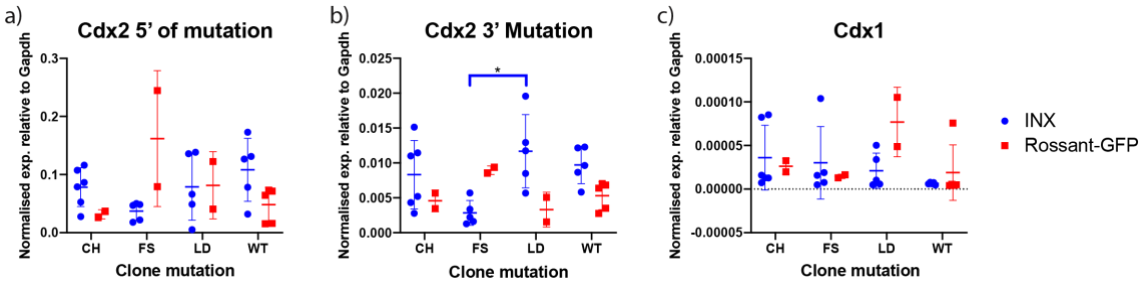



**Figure 6.8: Relative expression of *Cdx2* in *Cdx2*-mutant clones is similar to wild-type expression.** qRT-PCR analysis of *Cdx2* 5' and 3' of exon 2 and *Cdx1* in compound heterozygous (CH), frameshift (FS) and large deletion (LD) *Cdx2*-mutant clones and 5 wild-type (WT) control clones, outlined in Table 6.2. Each data point represents a single clone. Gene expression was normalised to *Gapdh*. Significance determined by two-way ANOVA test: * $p \leq 0.0332$. Error bars show SD.

*Cdx2* expression in edited clones was largely consistent with that observed for wild-type clones (Figure 6.8a-b). Given the absence of nonsense mediated decay for *Cdx2* transcripts in mutant clones, it was unlikely that transcriptional adaptation would occur (Lindeboom et al., 2016). In keeping with this, *Cdx1* expression was negligible and unchanged between wild-type and mutant clones (Figure 6.8c) and no *Cdx4* transcripts were detected in either wild-type or mutant clones (data not shown). It is possible, although unlikely, that other untested homeobox transcription factors may be upregulated to compensate for the absence of *Cdx2* instead.

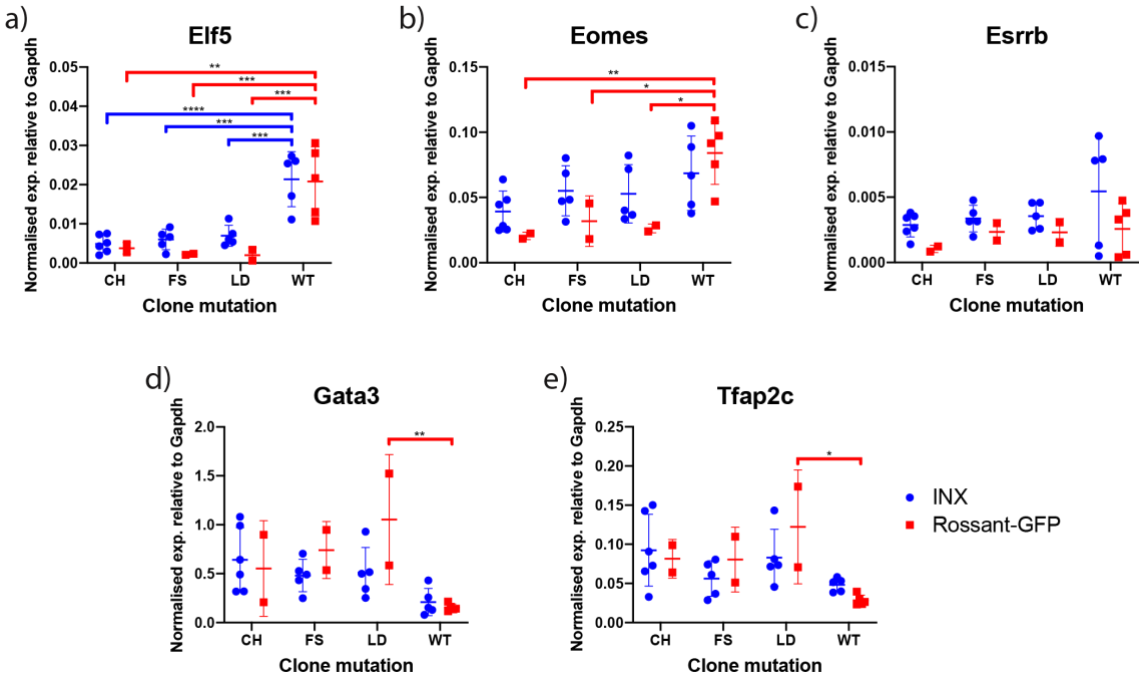



**Figure 6.9: Trophoblast stem cell marker gene expression in *Cdx2*-mutants.** qRT-PCR analysis for TSC marker genes in compound heterozygous (CH), frameshift (FS) and large deletion (LD) *Cdx2*-mutant clones and 5 wild-type (WT) control clones, outlined in Table 6.2. Each data point represents a single clone. Gene expression was normalised to *Gapdh*. Significance determined by two-way ANOVA test: * p≤0.0332, ** p≤0.0021, *** p≤0.0002, **** p≤0.0001. Error bars show SD.

Next, I asked whether the expression of other trophoblast stem cell markers was altered. *Elf5* was significantly downregulated in all mutant clones over WT controls (Figure 6.9a). All Rossant-GFP mutants, irrespective of the mutation type, showed a significant reduction in *Eomes* expression, but INX mutant clones did not (Figure 6.9b). *Esrrb* expression was unchanged in all mutant clones (Figure 6.9c). *Gata3* and *Tfap2c* are upregulated after both short-term (Chapter 4.3, Figure 4.4c) and long-term *Cdx2* knockdown (Chapter 5.1, Figure 5.2) and, although not significant in most cases, the same is true for most *Cdx2* mutant clones, irrespective of their mutation (Figure 6.9c,d). In sum, several core TSC marker genes change in expression, but only *Elf5* expression is universally affected in all clones.

To confirm that *Cdx2* mutations drive TSCs to differentiate into parietal trophoblast giant cells (p-TGCs), as in *Cdx2* knockdown, I examined the expression of TGC markers (Figure 6.10). The pan-TGC marker *Hand1* and mature TGC markers *Prl3d1, Prl3d3* and *Prl2c2* were upregulated in most mutant clones. However, this was not statistically significant because their expression was highly variable between mutant clones (Figure 6.10a-d), probably caused by the diverse range of phenotypes

between clones (Table 6.2). *Ctsq*, a marker of sinusoidal TGCs (s-TGCs) and channel TGCs (Ch-TGCs), was significantly upregulated in both 'large deletion' Rossant-GFP clones, but its expression was low in comparison to other TGC marker genes (Figure 6.10e). Therefore, the upregulation of *Ctsq* in these clones may not be biologically relevant. Overall, the co-expression of all p-TGC marker genes and absence of *Ctsq* expression indicates that *Cdx2*-mutants differentiate into p-TGCs.

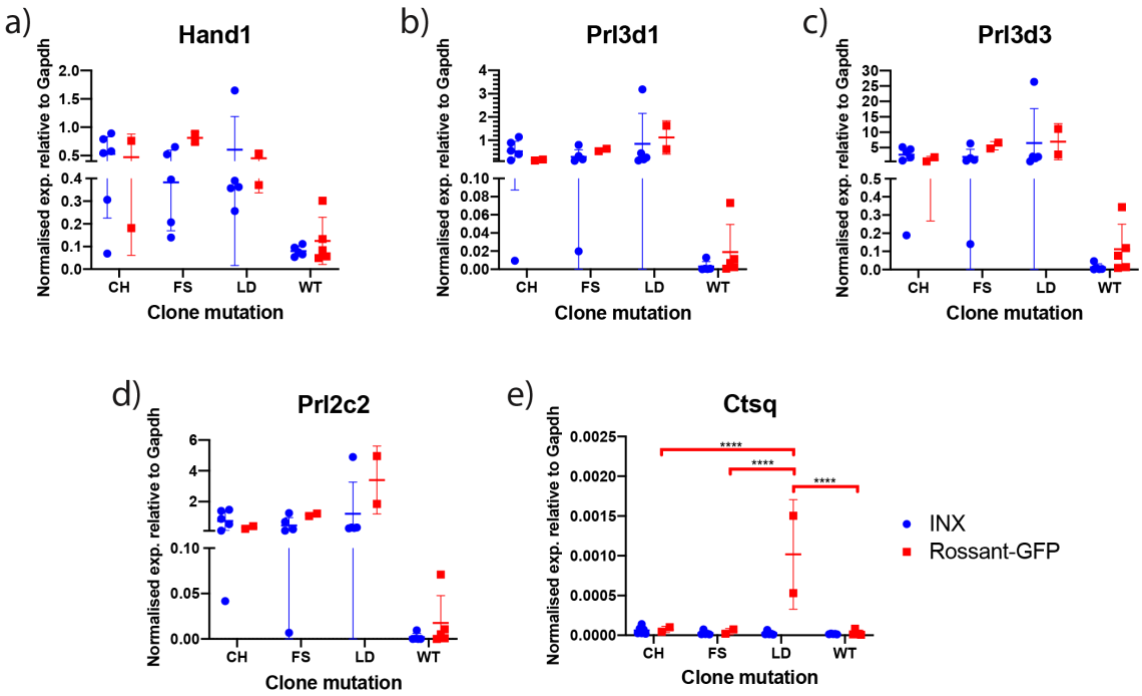



**Figure 6.10: Parietal trophoblast giant cell markers are upregulated in *Cdx2*-mutant clones.** qRT-PCR analysis for TGC marker genes in compound heterozygous (CH), frameshift (FS) and large deletion (LD) *Cdx2*-mutant clones and wild-type (WT) control clones. Co-expression of *Prl3d1*, *Prl3d3*, and *Prl2c2* and absence of *Ctsq* expression marks p-TGCs. Each data point represents a single clone. INX clones are shown in blue and Rossant-GFP are shown in red. Gene expression was normalised to *Gapdh*. Significance determined by two-way ANOVA test: **** p≤0.0001. Error bars show SD.

Next, I asked whether the expression levels of other trophoblast lineage markers were dysregulated in *Cdx2* mutant clones. Both the syncytiotrophoblast-I (Syn-I) progenitor marker *Snai1*, and Syn-I marker *Syna*, were upregulated in some, but not all, *Cdx2* mutant clones (Figure 6.11a-b). This upregulation was more consistent and statistically significant in Rossant-GFP mutant clones than in INX mutant clones. Both *Syna* and *Snai1* expression increased to a level comparable to that observed in *Cdx2* knockdown TSCs and in growth factor withdrawal differentiated cells (Figure 6.11, Figure 5.2). This data suggests that mutant clones may have acquired some Syn-I fate, but the TGC phenotype was more prominent.

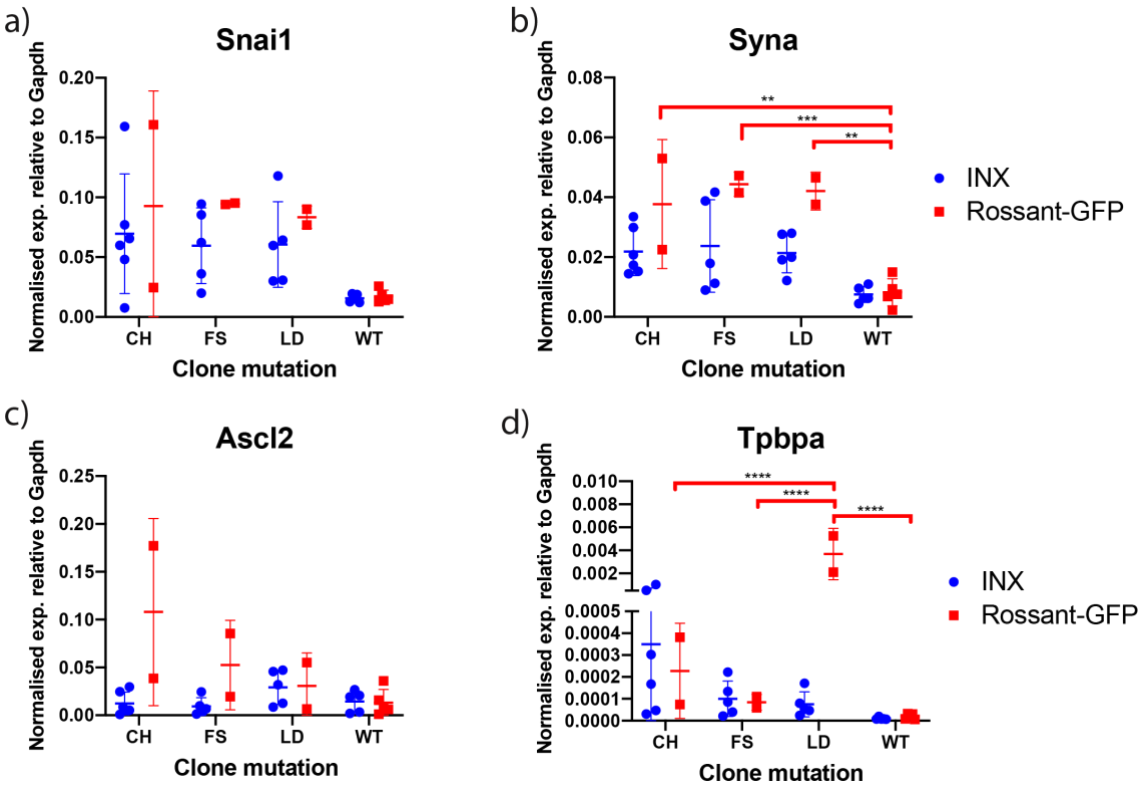



**Figure 6.11: *Cdx2*-mutant clones do not progress through a progenitor fate as they differentiate.** qRT-PCR analysis of Syncytiotrophoblast-I (Syn-I) marker genes *Snai1* and *Syna*, progenitor marker *Ascl2* and spongiotrophoblast and progenitor marker *Tpbpa*. Analysis was performed in compound heterozygous (CH), frameshift (FS) and large deletion (LD) *Cdx2*-mutant clones and 5 wild-type (WT) control clones, outlined in Table 6.2. Each data point represents a single clone. Gene expression was normalised to *Gapdh*. Significance determined by two-way ANOVA: ** p≤0.0021, *** p≤0.0002, **** p≤0.0001. Error bars show SD.

In Chapter 5, I showed that the progenitor marker genes *Ascl2* and *Tpbpa* were not upregulated after short-term nor long-term *Cdx2* knockdown (Chapter 3, Figure 5.2, Figure 5.8c). This suggests that *Cdx2* knockdown cells do not progress through a typical progenitor fate as they differentiate. In contrast to *Cdx2* knockdown, several Rossant-GFP *Cdx2* mutant clones, upregulated *Ascl2* over WT levels (Figure 6.11c). *Tpbpa,* also a marker of the spongiotrophoblast lineage, was not expressed in WT clones of either cell line. Almost all *Cdx2* mutant clones upregulated *Tpbpa*, but this was low and not statistically significant in most cases. However, both Rossant-GFP large deletion (LD) mutant clones significantly upregulated *Tpbpa* expression (Figure 6.11d). These Rossant-GFP LD clones were also those that uniquely upregulated *Ctsq* expression (Figure 6.10e) indicating that there may be something unique about these clones. In all, given the low expression of *Tpbpa* and *Ascl2* in most *Cdx2*-mutant clones and the high levels of differentiation, it can be concluded that TSCs lacking *Cdx2* do not progress through a progenitor fate during TGC differentiation.

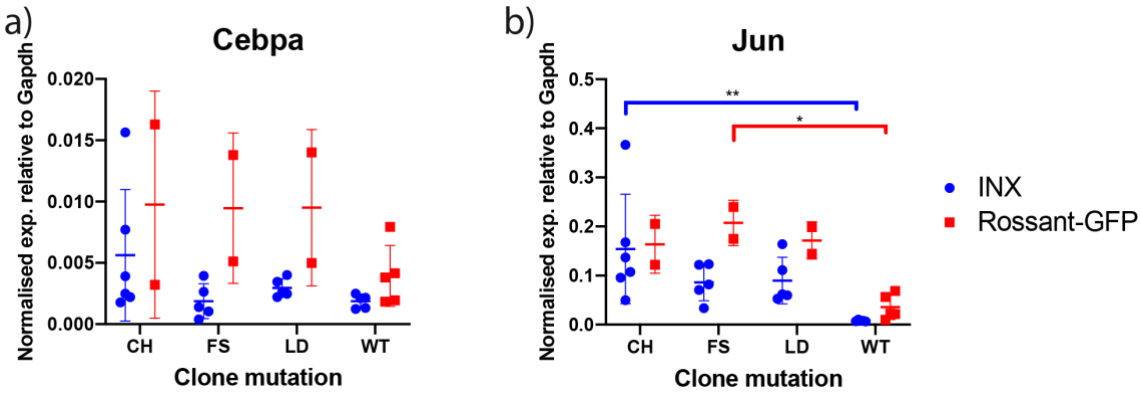



**Figure 6.12: Expression of potential TFAP2C co-factors in *Cdx2* mutants.** qRT-PCR analysis was performed in compound heterozygous (CH), frameshift (FS) and large deletion (LD) *Cdx2* mutant clones and 5 wild-type (WT) control clones, outlined in Table 6.2. Each data point represents a single clone. Significance determined by two-way ANOVA: * p≤0.0332, ** p≤0.0021. Error bars show SD.

Sites of increased accessibility, which was attributed to TFAP2C binding, are different between short-term *Cdx2* knockdown and growth factor withdrawal induced differentiation (-GF). Motif analysis attributed this to different transcription factor and/or co-factor binding. Two such motifs, the Jun-AP1 and CEBP motifs, are uniquely enriched in *Cdx2* knockdown and growth factor withdrawal conditions, respectively (Chapter 5.2, Figure 5.5). *Cebpa*, also a marker of the syncytiotrophoblast-II (Syn-II) lineage, was upregulated in the absence of growth factors but downregulated in transient *Cdx2* knockdown (Chapter 5.3, Figure 5.7) (Simmons et al., 2008). Few INX *Cdx2* mutant clones expressed *Cebpa* beyond the range observed in WT cells (Figure 6.12a). In Rossant-GFP *Cdx2* mutants, *Cebpa* was upregulated in one clone from each genotype, but its expression is also more variable in Rossant-GFP WT clones (Figure 6.12a). In contrast, *Jun* is upregulated after both short-term *Cdx2* knockdown and differentiation by growth factor withdrawal (Chapter 5.3, Figure 5.8). All *Cdx2* mutant clones upregulated *Jun* expression over wild-type clones, but this was not significant (Figure 6.12b). Further work is required to determine whether other transcription factors, whose motifs are differentially enriched in sites of gained accessibility between *Cdx2* knockdown and growth factor withdrawal conditions, are also differentially expressed in *Cdx2* mutant clones.

## 6.4 CRISPR clones reveal the importance of *Cdx2* to the trophoblast stem cell gene regulatory network

The work described above shows that the transcriptional profiles of *Cdx2* mutant clones and long-term *Cdx2* knockdown are similar. In both cases, the TSC marker genes *Elf5* and *Eomes* are downregulated, whereas *Gata3*, *Tfap2c* and several TGC markers are upregulated. However, the expression of these dysregulated genes varied between *Cdx2* mutant clones, perhaps reflecting their heterogenous morphological appearance (Table 6.2). I next explored whether there are any correlations between the changes in expression of analysed genes that could explain why some clones continue to proliferate whilst others differentiate (Figure 6.13).

Those *Cdx2* mutant clones that expressed small amounts of *Cdx1* are not those that downregulated *Cdx2*, nor those that expressed alternative splicing variants of *Cdx2* (underlined, Figure 6.13). Therefore, the continued survival of *Cdx2* mutant clones cannot be explained by genetic compensation from another *Cdx* family member.

Rossant-GFP *Cdx2* mutant clones that most significantly downregulated *Elf5* tended to downregulate *Eomes* more, downregulate *Esrrb* less and upregulate *Gata3* and *Tfap2c* the most (Figure 6.13). This trend is either not present or is less obvious in INX clones, perhaps representing biological difference between these TSC lines. Furthermore, those mutant clones that downregulated *Elf5*, *Eomes*, and *Esrrb* the most also showed the strongest upregulation of markers of differentiation, including p-TGC markers *Prl3d1, Prl2c2* and *Prl3d3*, the spongiotrophoblast marker *Tpbpa* and Syn-I associated markers *Syna* and *Snai1* (Figure 6.13). *Jun* upregulation is also predictive of TGC marker upregulation (Figure 6.13), perhaps indicating that its expression required during TGC differentiation as predicted in Chapter 5.

Given that most TSC marker genes are not expressed in differentiating cells, this observed correlation in gene expression is not surprising. *Cdx2* mutant clones are, therefore, heterogeneous populations in which the expression of TSC marker genes and differentiation marker genes changes depending on the rate at which they are differentiating.

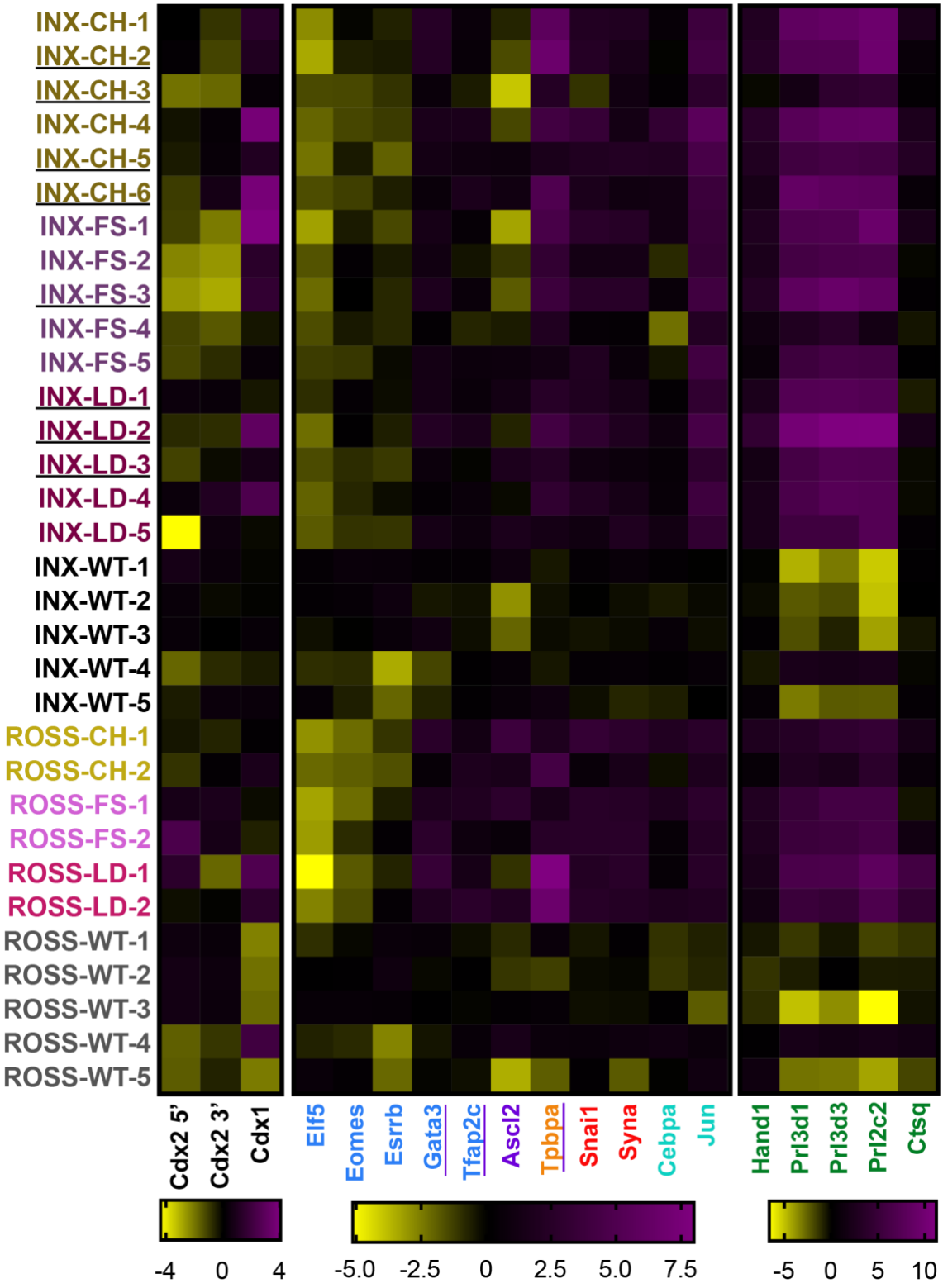



**Figure 6.13: Patterns in differential gene expression between *Cdx2*-mutant and WT clones may explain the observed heterogeneity in their differentiation.** qRT-PCR data for core trophoblast genes displayed as $Log_2$ fold change in normalised expression over average expression in the corresponding cell line WT clones. Those trophoblast lineages assayed are trophoblast stem cells (blue), progenitors (purple), spongiotrophoblast (orange), syncytiotrophoblast-I (red) and trophoblast giant cell (green). Motifs for *Cebpa* and *Jun* are differentially enriched at sites of gained accessibility after short-term *Cdx2* knockdown and differentiation by growth factor withdrawal. $Log_2$ fold change is displayed on different scales for different marker groups.

As the TSC markers *Elf5*, *Eomes* and *Esrrb* are not expressed in differentiated TSCs, it is difficult to know whether their downregulation in *Cdx2* mutant clones occurred in all cells or just those that were differentiating. To examine this, immunofluorescence was performed for multiple TSC marker genes on an exemplar *Cdx2* mutant clone (INX-FS-5) and a wild-type clone (INX-WT-2) (Figure 6.14).

The expression of CDX2 was determined in these clones using two antibodies in immunolabelling assays. Clone INX-FS-5 is homozygous for a 31bp deletion and does not express an alternatively spliced variant (Table 6.2). The first CDX2 antibody, used to generate ChIP-seq libraries in Chapter 4 (A300-691A, Bethyl), recognises an epitope of CDX2 between residues 250 and 300 of human CDX2. As this epitope is downstream of the 31bp deletion in INX-FS-5, it would be unlikely to recognise the frameshift downstream of this deletion. Low-level signal was observed using the CDX2 Bethyl antibody (Figure 6.14). Although predicted to recognise a specific epitope, the CDX2 Bethyl antibody is a polyclonal antibody, and may recognise other epitopes with lower efficiency. The epitope recognised by the second CDX2 antibody (CDX2-88, Abcam) is not disclosed by any distributor. No signal was seen when using this CDX2-88 antibody (Figure 6.14).

To mark differentiated cells, clones were stained for the TGC marker HAND1. Like all other TGC marker genes, *Hand1* expression was highly variable between clones (Figure 6.10a). No *Cdx2* mutant clone upregulated *Hand1* more than 20-fold over WT controls and only half upregulated *Hand1* over 5-fold (Figure 6.10a). Despite the modest increase in *Hand1* expression (1.7-fold more than in WT control cells), INX-FS-5 contained many cells resembling TGCs: HAND1 expressing cells with large nuclei and prominent actin stress fibres. In contrast, the wild-type control, INX-WT-2, did not express HAND1, nor did this clone possess any other phenotypic hallmarks of TGCs (Figure 6.14).

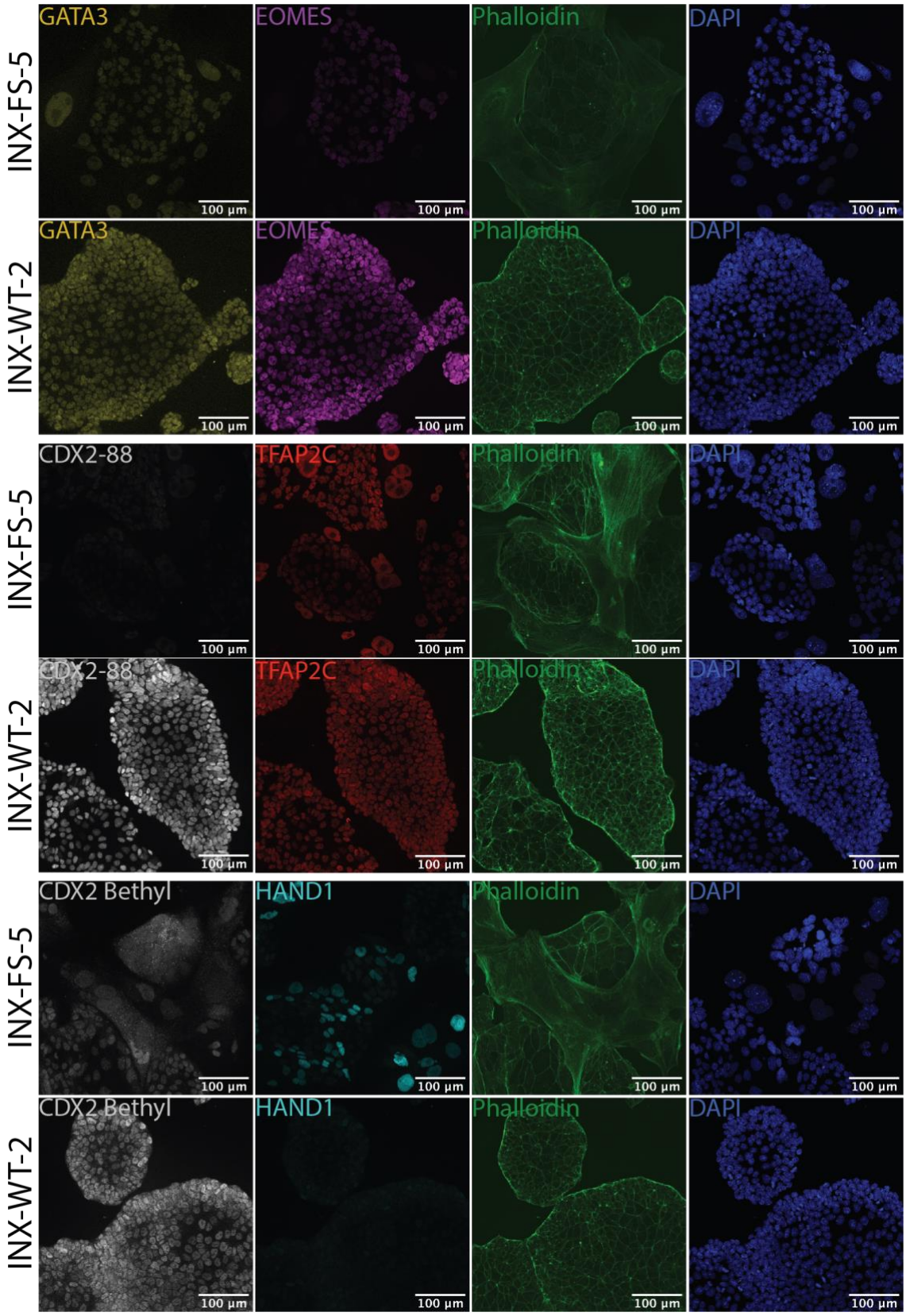



**Figure 6.14: TSC marker gene expression is significantly perturbed at the protein level in *Cdx2*-mutants.** Immunofluorescence was performed for TSC marker genes (GATA3, CDX2, EOMES, TFAP2C) and the TGC marker gene HAND1 in a *Cdx2*-mutant clone (INX-FS-5) and wild-type control (INX-WT-2). Phalloidin staining was used to assess the presence of actin stress fibres, a phenotypic hallmark of TGCs.

Interestingly, the expression of TSC markers at the transcript level does not reflect their protein expression. In contrast to its 2.3-fold upregulation at the transcript level in INX-FS-5, GATA3 is downregulated in all cells relative to INX-WT-2 (Figure 6.13, Figure 6.14). *Tfap2c* is upregulated 1.7-fold in INX-FS-5 over WT cells (Figure 6.13). *Cdx2* mutant cells that resemble TGCs expressed more TFAP2C than other cells, indicating that it is upregulated during the differentiation of *Cdx2* mutant cells as in *Cdx2* knockdown conditions. However, TFAP2C is lower in all other cells in this *Cdx2* mutant population than in INX-WT-2 cells, suggesting its expression is lower in proliferating *Cdx2* mutant cells than wild-type cells (Figure 6.14). *Eomes mRNA* expression in most INX mutant clones was within the range of expression observed in WT control cells (Figure 6.9b). However, EOMES is almost completely lost in all INX-FS-5 cells (Figure 6.14). In all, the data suggests that the expression of core TSC genes at the protein level are significantly affected when *Cdx2* is targeted.

## 6.5 Summary

CRISPR-Cas9 gene targeting at the *Cdx2* locus using two gRNAs in combination was efficient and reproducible, generating consistent mutations between experiments and cell lines. *Cdx2* mutant alleles were outcompeted by WT alleles in polyclonal populations, probably due to an increase in the number of wild-type cells at the expense of clones containing mutations. Unexpectedly, given the phenotype in long-term *Cdx2* knockdown, it was possible to maintain *Cdx2*-mutant clonal lines in culture for several passages. Despite this, the outcome of *Cdx2* knockdown and knockout are comparable, with both driving differentiation into p-TGCs without progressing through a progenitor fate in both the INX and Rossant-GFP cell lines.

It is possible to predict which clones contain *Cdx2* mutations by their phenotypic appearance, with most mutant clones showing sporadic differentiation within the first few passages. Of those clones that were successfully genotyped, 59% of clonal lines contained two or more *Cdx2* mutant alleles in different ratios, suggesting that some populations were either not clonal or not diploid. Although normal karyotypes have been found in TSC lines at early passages, aneuploidy is prevalent in TSC lines over passage 40 (Tanaka et al., 1998, Kubaczka et al., 2014), as are both the INX and Rossant-GFP TSC lines used in this thesis (personal correspondence with Daniel Snell). *Cdx2* mutant populations also differentiate into TGCs and are, therefore, undergoing endoreduplication. Proportions of the genome are commonly under- or over-replicated as a result of endoreduplication, although such copy number variants have not been reported for the *Cdx2* locus (Hannibal and Baker, 2016, Hannibal et al., 2014). Whether clonal or otherwise, the mutants used for subsequent qRT-PCR and IF analyses contained no WT *Cdx2* alleles.

The two most common mutations in polyclonal populations were also found in 71% of all successfully genotyped clones: a 42bp deletion and a 31bp deletion. The 42bp deletion would result in the loss of 14 amino acids from the conserved homeobox domain, namely the second helix and loop immediately adjacent to the DNA binding third helix. In contrast, the 31bp deletion results in a frameshift and is predicted to generate a truncated protein 42 amino acids smaller than wild-type CDX2 that is non-functional. All of the clones validated express *Cdx2* transcripts to similar extents as

WT clones. Immunofluorescence on a 31bp deletion mutant using two different anti-CDX2 antibodies validated that CDX2 expression is significantly reduced or absent and that 31bp deletion homozygous mutant clones are *Cdx2* knockout TSC lines.

Given that all validated *Cdx2* mutant clones largely contain these two mutations, the heterogeneity in the rate at which they differentiate is surprising. Rossant-GFP *Cdx2* mutant clones cannot be maintained for the same number of passages as INX for the same mutation. This suggests that they cannot compensate for loss of *Cdx2* to the same extent. One observed difference between the cell lines is that half of all INX *Cdx2*-mutant clones express at least one alternatively spliced variant of *Cdx2*, whereas no Rossant-GFP clones do. Alternative splicing may occur due to disruption of exonic splicing enhancers in the regions deleted in mutants. Yet, why this only occurs in INX clones and what the function and/or significance of this observation is remains unclear.

Another potential explanation is that *Eomes* expression was shown to be more consistently and significantly downregulated in Rossant-GFP clones than INX clones. A core TSC marker gene, *Eomes* is known to work in tandem with *Elf5* and *Tfap2c* to maintain a subsection of TSC gene regulatory network (Latos et al., 2015b). In combination with the significant downregulation in *Elf5* observed in both cell lines, this difference in *Eomes* expression may explain why Rossant-GFP clones are inherently less capable of maintaining proliferation in the absence of *Cdx2*. Despite this apparent difference in *Eomes* transcript expression between INX and Rossant-GFP clones, EOMES protein staining was negligible in a homozygous 31bp deletion INX clone.

Both *Gata3* and *Tfap2c* are upregulated in most mutant clones, in agreement with published literature showing that they are upregulated in differentiation (Latos et al., 2015b, Ralston et al., 2010). Although *Gata3* transcript increased, GATA3 protein was significantly reduced in an INX *Cdx2*-mutant clone. TFAP2C staining was lower in the proliferating cells in the INX *Cdx2*-mutant clone, but its expression is similar or higher in cells displaying morphological characteristics of differentiation. This may indicate an essential role of TFAP2C in driving differentiation towards TGCs as

observed in *Cdx2* knockdown. In all, this shows that the GRN of *Cdx2* mutant clones is significantly affected.

Although *Elf5* and *Eomes* may not be immediately downstream of *Cdx2*, given that their expression is unaffected in transient *Cdx2* knockdown, the loss of *Cdx2* has a significant effect on their expression. However, given the heterogeneity of transcript expression levels and morphological appearance of clones, further immunostainings on other mutant clones are required to confirm that differences in the GRN are universal to TSCs when *Cdx2* is deleted.

Given the described changes to multiple TSC marker genes described above, how *Cdx2*-mutant cells compensate and continue to self-renew for several passages is unclear. One potential hypothesis tested was that transcriptional adaptation, known to occur when mutant transcripts are expressed and rapidly undergo nonsense mediated decay, might be important in driving compensation (El-Brolosy et al., 2019). As *Cdx2* transcripts do not undergo nonsense mediated decay, it was unsurprising that *Cdx1* and *Cdx4* expression was negligible or absent. Whilst there was no link between reduction in *Cdx2* transcript expression and an upregulation of *Cdx1*, there are other genes with similar sequence homology to *Cdx2* that were not tested. However, genetic compensation is unlikely to be the reason for the prolonged survival of *Cdx2* mutant clones.

Due to the high heterogeneity in differentiation phenotype, I have not been able to conclude how *Cdx2*-mutant clones compensate for the loss of *Cdx2* and subsequent changes to the TSC GRN. To more accurately assess this, an unbiased screen of the entire transcriptome is required.

# Chapter 7. Discussion

In this thesis I have examined the short- and long-term consequences of reduced *Cdx2* expression in trophoblast stem cells. I have also addressed the role of *Cdx2* within the established trophoblast gene regulatory network. The key finding of my work is that *Cdx2* plays a significant role in maintaining the balance of the TSC gene regulatory network and that it prevents direct TSC differentiation into parietal trophoblast giant cells. *Cdx2* is, therefore, a gatekeeper of mouse TSC potency.

**_Cdx2_ and the trophoblast stem cell gene regulatory network**

Transient *Cdx2* knockdown in trophoblast stem cells, to about 60% of control levels, perturbs the chromatin landscape and transcriptome but has no effect on the expression of most TSC marker genes. To some extent, this lack of effect is unsurprising, because even in normal TSCs, expression of *Cdx2*, as well as that of *Eomes* and *Gata3*, declines with time in culture. One possibility is that *Cdx2* knockdown causes TSCs to adopt a more differentiated cell type only transiently (Frias-Aldeguer et al., 2019).

In contrast, long-term reduction or loss of *Cdx2* expression affects the expression of most TSC marker genes (Latos et al., 2015a) and drives differentiation into parietal trophoblast giant cells. Both *Elf5* and *Eomes* are downregulated after prolonged *Cdx2* loss, while *Tfap2c* and *Gata3* are upregulated. These results suggest that *Cdx2* is important for long-term maintenance of the TSC gene regulatory network.

Despite their upregulation at the transcript level, TFAP2C and GATA3 proteins are expressed at lower levels in proliferating cells of surviving *Cdx2* mutant clones than in their wild-type counterparts. What allows *Cdx2* mutant clones to survive? It is possible that they adopt a distal extraembryonic ectoderm fate (*Cdx2* is not expressed in distal extraembryonic ectoderm), but the lack of *Elf5* and *Eomes* expression in *Cdx2* mutant clones suggests this is unlikely (Donnison et al., 2015). Another possibility is based on the observation that *Esrrb* expression fails to be downregulated in these surviving mutant clones. ESRRB blocks the expression of differentiation markers (Latos et al., 2015a), and this may enable *Cdx2* mutant clones to self-renew. More work is required to explore the role of ESRRB in *Cdx2* mutants.

**Not all *Cdx2* mutant clones are the same**

TSCs are heterogeneous and can adopt a range of trophoblast stem cell fates (Frias-Aldeguer et al., 2019). *Cdx2* mutant clones, even those with the same mutations in the same TSC line, differ with respect to their ability to self-renew or differentiate. The behaviour of a *Cdx2* clone may depend on the type of TSC that has been mutated, with one trophoblast fate responding differently to *Cdx2* loss than another. If the proportions of TSC subpopulations are different between different cell lines, this may provide an explanation of why clones from the INX cell line survive longer than Rossant-GFP clones. To address this, RNA-seq analysis might be required to provide a global insight into clone heterogeneity and their compensatory mechanisms for the loss of the TSC GRN.

**Trophoblast giant cells differentiated by *Cdx2* loss are not derived from a progenitor population**

When they move away from growth factor signals in the extraembryonic ectoderm, TSCs differentiate as spongiotrophoblast and TGCs via progenitors in the ectoplacental cone (EPC) that are marked by *Ascl2* and *Tpbpa* (Guzman-Ayala et al., 2004, Hughes et al., 2004, Latos and Hemberger, 2016). However, 50% of secondary parietal trophoblast giant cells come from non-*Tpbpa* expressing cells (Simmons et al., 2007) and embryos lacking *Ascl2* are nevertheless able to form p-TGCs (Guillemot et al., 1994). This indicates that p-TGCs can form *in vivo* without progressing through an EPC fate.

*Ascl2* and *Tpbpa* are not significantly upregulated when *Cdx2* expression is reduced or abolished, suggesting that the TSCs that differentiate as a result of *Cdx2* loss do not progress through an EPC progenitor fate. The absence of *Ascl2* expression is of particularly significance because *Ascl2* antagonises the pan-TGC marker *Hand1* (Scott et al., 2000), which itself is upregulated after transient *Cdx2* knockdown. When overexpressed in TSCs, *Hand1* causes about 30% of cells to differentiate towards TGCs (Scott et al., 2000), but this differentiation is significantly less efficient than the homogeneous differentiation observed in long-term *Cdx2* knockdown. Furthermore, HAND1 is not expressed in proliferating *Cdx2* mutant cells, indicating that *Hand1* upregulation is not the driver of direct TGC differentiation when *Cdx2* expression is reduced. Instead, I propose that a relationship between CDX2 and TFAP2C, direct

or otherwise, is essential to modulate this direct differentiation into p-TGC. *Cdx2* knockdown in TSCs provides an *in vitro* model system to characterise this relationship and direct p-TGC differentiation.

***Tfap2c*: the multipurpose trophoblast transcription factor**

TFAP2C is a versatile transcription factor that has distinct roles at different stages of trophoblast development (Kuckenberg et al., 2012, Choi et al., 2012, Cao et al., 2015, Latos et al., 2015b, Nelson et al., 2017). It is a core TSC transcription factor that is expressed in stoichiometric ratios with *Eomes* and *Elf5* in TSC. In TSCs, interaction is stronger between ELF5 and EOMES proteins, but all three transcription factors bind together around TSC marker genes (Latos et al., 2015b). During differentiation by growth factor withdrawal *in vitro*, *Tfap2c* and *Elf5* are upregulated, whilst *Eomes* is downregulated (Latos et al., 2015b). As a result, during differentiation ELF5 preferentially interacts with TFAP2C, globally increasing chromatin accessibility by binding near differentiation-associated genes (Latos et al., 2015b, Nelson et al., 2017). Interestingly, CDX2 also directly interacts with EOMES in TSCs (Latos et al., 2015a), but how relevant this is to the balance between stemness and differentiation is unclear.

As is observed following growth factor withdrawal, chromatin accessibility increased after transient *Cdx2* knockdown. Increased accessibility in both cases is attributed to TFAP2C, but the sites themselves differ between *Cdx2* knockdown and growth factor withdrawal differentiation. Motif analysis suggested that these differences are driven by different co-factors. In keeping with this, although most of the genes dysregulated by *Cdx2* knockdown are affected by growth factor withdrawal, 31% of these genes changed in opposite directions. Interestingly, neither of the known TFAP2C interactors, *Elf5* and *Eomes*, were dysregulated in transient *Cdx2* knockdown cells and their motifs were not identified in sites of gained accessibility. Instead, *Cebpa* and *Jun* were identified as candidate TFAP2C co-factors under growth factor withdrawal and *Cdx2* knockdown conditions, respectively. In all, TFAP2C appears to be repurposed depending on what co-factors are available, either maintaining undifferentiated cells or driving differentiation into different fates.

I confirmed that half of those sites that gained accessibility upstream of the *Hand1* locus in *Cdx2* knockdown cells are functional enhancers for the *Hand1* promoter. TFAP2C binds to these sites, as shown by published TFAP2C ChIP-seq data (Latos et al., 2015b), but its functional requirement for the activity of these enhancers requires further validation. As several of these enhancers behave differently in the presence and absence of growth factors, context is essential for regulating TFAP2C behaviour. Therefore, I hypothesise that both the transcriptional changes that occur after *Cdx2* knockdown and the change in the balance of CDX2 and TFAP2C expression alters the behaviour of TFAP2C. As a result, reduced *Cdx2* expression drives the observed direct differentiation towards p-TGCs. To understand why TFAP2C behaviour changes in different trophoblast contexts, one could use mass spectrometry to determine the factors that TFAP2C interacts with under stem cell, *Cdx2* knockdown and growth factor withdrawal differentiation conditions.

**CDX2 is repurposed from pre- to post-trophoblast establishment**

*Cdx2*-null embryos are incapable of forming trophoblast outgrowths (Strumpf et al., 2005), confirming that *Cdx2* is essential for reinforcing commitment to the trophoblast fate. As shown in this thesis, *Cdx2* is essential for maintaining undifferentiated TSCs, suggesting that *Cdx2* also plays a key role in the established trophoblast fate.

By combining CDX2 ChIP-seq data with ATAC-seq data in TSCs, I showed that most CDX2 binding sites are located in regions of open chromatin in TSCs, suggesting that CDX2 is not a pioneer factor. There are, however, a small number of CDX2 binding sites which are inaccessible in TSCs. Bearing this in mind, it is possible that CDX2 is a pioneer factor in TSCs, although this is unlikely given that TSCs are a stable cell type and *Cdx2* is downregulated as TSCs differentiate.

To address the possibility that CDX2 plays different roles pre- and post-trophoblast commitment, I separated CDX2 binding sites in established TSCs based on their accessibility in mESCs. CDX2 binding sites that are ‘accessible’ in mESCs were found around genes expressed in early development and are highly enriched for multiple TEAD family of motifs as well as SOX, ESRRB and homeobox. TEAD4 regulates *Cdx2* expression during trophectoderm establishment (Nishioka et al., 2009, Rayon et al., 2014) and directly interacts with CDX2 in TSCs (Latos et al.,

2015a). Since mESC-accessible sites are unaffected by *Cdx2* knockdown, CDX2 binding here is likely recruited by TEAD4. I, therefore, hypothesise that these sites represent binding targets of CDX2 during trophectoderm establishment that remain bound in TSCs.

CDX2 binding sites that were inaccessible in mESCs were located around genes expressed in later embryonic development. Although they are highly enriched for the homeobox motif, the accessibility of most of these sites is unaffected by *Cdx2* knockdown. This is largely due to the enrichment of other TSC maker gene motifs at sites that are accessible in TSCs and the enrichment of homeobox motifs at poorly accessible or inaccessible sites in TSCs. In all, by separating CDX2 binding sites based on their accessibility in mESCs and TSCs, I have shown that CDX2 binding is highly opportunistic or recruited by other factors, perhaps explaining why so few CDX2 binding sites lose accessibility in *Cdx2* knockdown. Therefore, while a subset of CDX2 binding sites may be critical to maintain stemness in TSCs, more work is required to understand the relevance of CDX2 binding at most of its binding sites.

Further work is also required to understand whether CDX2 has different binding sites at different stages of trophoblast lineage development *in vivo* that can explain its different role pre- and post-trophoblast commitment. Although there are no ATAC-seq libraries available for the trophectoderm and the extraembryonic ectoderm, these would be invaluable if we are to better understand the trophoblast lineage. Using these libraries, differentially accessible chromatin regions that overlap with CDX2 binding sites could be validated using previously established ChIP-qPCR protocols for mouse blastocysts (Home et al., 2009).

**TSC-like cells are not real TSCs**

Whether transdifferentiation of mESCs generates *bona fide* TSCs is contentious because the resulting TSC-like cells have distinct transcriptomes, DNA methylation patterns and chromatin landscape profiles from *bona fide* TSCs (Cambuli et al., 2014, Rhee et al., 2017). *Bona fide* TSC have highly similar chromatin accessibility, even when different cell lines and analysis pipelines are used (Nelson et al., 2017). Therefore, comparing the chromatin landscape is a robust approach to validate TSC identity.

In this thesis, I re-analysed published CDX2 ChIP-seq libraries from TSC-like cells made by repressing *Oct4* in mESCs (Adachi et al., 2013) and compared them to libraries I made from embryo outgrowth-derived TSCs. CDX2 binding in TSC-like cells was different to that in embryo-outgrowth derived TSCs. Furthermore, the unique CDX2 binding sites in TSC-like cells were inaccessible in both *bona fide* TSCs and in the mESC line from which TSC-like cells were derived from. Additionally, over 78% of CDX2 binding in TSC-like cells derived by overexpressing *Cdx2* in mESCs are unique from any other CDX2 ChIP-seq library analysed in this thesis (Rhee et al., 2017). Whilst it is possible that some CDX2 binding sites in TSC-like cells are uniquely bound in trophectoderm specification, the majority are likely biologically irrelevant. As such, caution must be taken in drawing conclusions from TSC-like cells.

**Conclusion**

In summary, my work confirms previous observations that loss of *Cdx2* expression in TSCs increases differentiation towards parietal trophoblast giant cells (Latos et al., 2015a, Niwa et al., 2005). However, in contrast to previous work, I show that this differentiation is homogenous, avoids the progenitor fate, and is mediated by the upregulation of *Tfap2c* and other co-factors. To the best of my knowledge, this is the first time *Tfap2c* and *Cdx2* have been associated with the direct differentiation of stem cells into parietal trophoblast giant cells. The direct differentiation of TSCs to TGCs had been previously shown *in vivo*, both from the mural trophectoderm and extraembryonic ectoderm, but is poorly understood. Here, I provide a potential mechanism for this phenomenon. As the behaviour of TFAP2C when *Cdx2* expression is perturbed is different to others contexts, I have provided new insight into the importance of the balance of CDX2 and TFAP2C expression and shown that this dynamic represents one of the main functions of CDX2 in TSCs. This is distinct from the role of CDX2 in repressing the pluripotency network during trophectoderm establishment, a role CDX2 probably exerts by its interaction with TEAD4. Finally, I showed that *Cdx2* mutant clones can be maintained *in vitro* for several passages but are heterogeneous and predisposed to differentiate into p-TGCs. *Cdx2* mutant clones may be used to further our understanding of the plasticity in the trophoblast gene regulatory network and better understand the role of *Cdx2* within this network.

# Appendix

## Appendix A: Cell Culture Media and Solutions

| Medium | Component | Source | Product Code | Volume | Final Conc. |
|---|---|---|---|---|---|
| TSCM | RPMI1640 +Glutamax | Thermo Fisher Scientific | 61870010 | 390mL | n/a |
| | Foetal bovine serum | Thermo Fisher Scientific | 16000044 | 100mL | 20% v/v |
| | Penicillin-streptomycin | Thermo Fisher Scientific | 150700A973163 | 5mL | 50 U/mL and 50 µg/mL |
| | Sodium pyruvate | Thermo Fisher Scientific | 11360070 | 5mL | 1mM |
| | 2-mercaptoethanol | Thermo Fisher Scientific | M6250 | 3uL | 0.1uM |
| TSCMM | Fresh TSCM | See above | n/a | 30% | n/a |
| | Conditioned TSCM | TSCM conditioned on MEFs | n/a | 70% | n/a |
| | Recombinant human FGF4 | Peprotech | 100-31 | | 1.5 µg/mL |
| | Heparin sodium salt | Sigma-Aldrich | H3149 | | 37.5 ng/mL |
| MEF | Advanced DMEM/F12 | Thermo Fisher Scientific | 12634028 | 440 ml | n/a |
| | Foetal bovine serum | Thermo Fisher Scientific | 16000044 | 50mL | 20% v/v |
| | Glutamax | Thermo Fisher Scientific | 35050038 | 5mL | 2mM |
| | Penicillin-streptomycin | Thermo Fisher Scientific | 150700A973163 | 5mL | 50 U/mL and 50 µg/mL |
| | 2-mercaptoethanol | Thermo Fisher Scientific | M6250 | 3uL | 0.1uM |

**Table A-1: Components of cell culture media used in this thesis**

| Solution | Component | Concentration | Chapter |
|---|---|---|---|
| KT lysis buffer | Tris-HCl (pH9.1) | 67 mM | 2.2.1 |
| | $(NH_4)_2SO_4$ | 15 mM | |
| | $MgCl_2$ | 3.5 mM | |
| | Bovine serum albumin | 150 µg/mL | |
| | Proteinase K | 800 µg/mL | |
| | Triton-X100 | 0.50% | |
| ChIP Cell lysis buffer | Tris-HCl (pH8.0) | 10 mM | 2.4.1.1 |
| | NaCl | 10 mM | |
| | IGEPAL CA-630 | 0.20% | |
| CEWB1 buffer | Tris-HCl (pH8.0) | 10 mM | 2.4.1.1 |
| | EDTA | 1 mM | |
| | Sodium dodecyl sulphate (SDS) | 0.10% | |
| | NaCl | 150 mM | |
| | Sodium deoxycholate | 0.25% | |
| | IGEPAL CA-630 | 1% | |
| RIPA buffer | HEPES (pH7.5) | 50 mM | 2.4.1.1 |
| | EDTA | 1 mM | |
| | LiCl | 500 mM | |
| | Sodium deoxycholate | 0.70% | |
| | IGEPAL CA-630 | 1% | |
| TEN buffer | Tris-HCl (pH8.0) | 10 mM | 2.4.1.1 |
| | EDTA | 1 mM | |
| | NaCl | 150 mM | |
| SDS Elution buffer | Tris-HCl (pH8.0) | 50 mM | 2.4.1.1 |
| | EDTA | 1 mM | |
| | Sodium dodecyl sulphate (SDS) | 1% | |
| ATAC Cell lysis buffer | Tris-HCl (pH7.4) | 10 mM | 2.4.2 |
| | NaCl | 10 mM | |
| | $MgCl_2$ | 3 mM | |
| | IGEPAL CA-630 | 0.10% | |

**Table A-2: Components of the solutions used in this thesis**

# Appendix B: Oligonucleotide Sequences

| Chapter | Oligo | Sequence | Application |
|---|---|---|---|
| 2 | mCIP_ F | CTCCCCAGGGGGATCCACGCCGCCACCATGGCCATCATCAAGGAG | Cloning |
| 2 | mCIP_R | GGTCTTAAAGGTACCTCAGGCACCGGGCTTGCGGG | Cloning |
| 2 | Hand1_Prom_209bp_F | GATCCTCGAGATAAAACTTGGGACCGCCAC | Cloning |
| 2 | Hand1_Prom_209bp_R | GATCAAGCTTAGGCTGGAGATGACACGAAG | Cloning |
| 2 | Hand1_Prom_283bp_F | GATCCTCGAGAGAGGGGAGGGACATAGGC | Cloning |
| 2 | Hand1_Prom_283bp_R | GATCAAGCTTAGGCTGGAGATGACACGAAG | Cloning |
| 2 | Hand1_Prom_673bp_F | GATCCTCGAGCTGAAGCAGGGGACCTAAAG | Cloning |
| 2 | Hand1_Prom_673bp_R | GATCAAGCTTAGGCTGGAGATGACACGAAG | Cloning |
| 2 | Hand1_Enh-A_F | GATCGTCGACATAATGGGTGGCCCTTTCC | Cloning |
| 2 | Hand1_Enh-A_R | GATCGTCGACGCCTCCTATCAACCAGGCTAC | Cloning |
| 2 | Hand1_Enh-B_F | TAAGGATCCGTCGACGCATTCTAAGGGATTTTGATGG | Cloning |
| 2 | Hand1_Enh-B_R | AGGGCATCGGTCGACTCAGGGCTGCGAAAACTC | Cloning |
| 2 | Hand1_Enh-C_F | GATCGTCGACTAGAGAGACCACCCCCTGAG | Cloning |
| 2 | Hand1_Enh-C_R | GATCGTCGACCAGTGGAATGCCATGCAG | Cloning |
| 2 | Hand1_Enh-D_F | TAAGGATCCGTCGACGCACTCCTGTCCTTTGTAGGTC | Cloning |
| 2 | Hand1_Enh-D_R | AGGGCATCGGTCGACGAGGAAGAACCTGAAATAAGC | Cloning |
| 2 | Hand1_Enh-E_F | GATCGTCGACAGGAAAGTGGGTTGGTGCT | Cloning |
| 2 | Hand1_Enh-E_R | GATCGTCGACCTTCTATCAGGCTGGGTTGG | Cloning |
| 2 | Hand1_Enh-F_F | GATCGTCGACTGCTTGCCAGAGTTGTGTTT | Cloning |
| 2 | Hand1_Enh-F_R | GATCGTCGACCCTCAGTGTCCCAAAATGGT | Cloning |
| 2 | Hand1_Enh-C1_F | GATCGTCGACGGAGAATCTTCTAGGTCTGTCTGG | Cloning |
| 2 | Hand1_Enh-C1_R | GATCGTCGACCGGGGAGAGAGGAGTAGAATG | Cloning |
| 2 | Hand1_Enh-C2_F | GATCGTCGACCAGAACACGGAAACGAGTTG | Cloning |
| 2 | Hand1_Enh-C2_R | GATCGTCGACTCGGGACCAGCTGTTAATTC | Cloning |
| 2 | sgRNA_G2a_F | CACCGTTAGTCGATACATCACCATC | Cloning |
| 2 | sgRNA_G2a_R | AAACGATGGTGATGTATCGACTAAC | Cloning |
| 2 | sgRNA_G2b_F | CACCGTACACAGACCATCAGCGGC | Cloning |
| 2 | sgRNA_G2b_R | AAACGCCGCTGATGGTCTGTGTAC | Cloning |
| 2 | sgRNA_G2c_F | CACCGCCACACTTGGGCTCTCCGAG | Cloning |
| 2 | sgRNA_G2c_R | AAACCTCGGAGAGCCCAAGTGTGGC | Cloning |
| 2 | sgRNA_G2d_F | CACCGCCTCTCGGAGAGCCCAAGTG | Cloning |
| 2 | sgRNA_G2d_R | AAACCACTTGGGCTCTCCGAGAGGC | Cloning |
| 2 | sgRNA_G2e_F | CACCGTCGATACATCACCATCAGG | Cloning |
| 2 | sgRNA_G2e_R | AAACCCTGATGGTGATGTATCGAC | Cloning |
| 2 | sgRNA_G2f_F | CACCGATGGTCTGTGTACACCACC | Cloning |
| 2 | sgRNA_G2f_R | AAACGGTGGTGTACACAGACCATC | Cloning |
| 2 | pGL2-Basic_F | GGCCTCTTCGCTATTACGC | Sanger Sequencing |
| 2 | LucNrev | CCTTATGCAGTTGCTCTCC | Sanger Sequencing |
| 2 | RVprimer4 | GACGATAGTCATGCCCCGCG | Sanger Sequencing |
| 2 | EnForKB | TGGGAGGTTTTTTAAAGCAA | Sanger Sequencing |
| 2 | hU6 | GACTATCATATGCTTACCGT | Sanger Sequencing |
| 2 | 5to3_Cdx2_F | GTCCCTCGCCACCATGTA | Sanger Sequencing |
| 2 | 5to3_Cdx2_R | TCCACTCATCTTCTGTCCTCAA | Sanger Sequencing |
| 3,4,5,6 | Cdx2_5'_F | AAACCTGTGCGAGTGGATG | qRT-PCR |
| 3,4,5,6 | Cdx2_5'_R | GCCGCTGATGGTCTGTGTA | qRT-PCR |
| 3,4,5,6 | Elf5_F | GGACTCCGTAACCCATAGCA | qRT-PCR |
| 3,4,5,6 | Elf5_R | TACTGGTCGCAGCAGAATTG | qRT-PCR |
| 3,4,5,6 | Eomes_F | GGAAGTGACAGAGGACGGTG | qRT-PCR |
| 3,4,5,6 | Eomes_R | GGAAGTGACAGAGGACGGTG | qRT-PCR |
| 3,4,5,6 | Esrrb_F | AACAGCCCCTACCTGAACCT | qRT-PCR |
| 3,4,5,6 | Esrrb_R | CTCATCTGGTCCCCAAGTGT | qRT-PCR |
| 3,4,5,6 | Gata3_F | GGGTTCGGATGTAAGTCGAG | qRT-PCR |
| 3,4,5,6 | Gata3_R | CCACAGTGGGGTAGAGGTTG | qRT-PCR |
| 3,4,5,6 | Tfap2c_F | TGCACACAGGGTATTGAAAC | qRT-PCR |
| 3,4,5,6 | Tfap2c_R | CGTCACCCACACAAACTAAA | qRT-PCR |
| 3,4,5,6 | Ascl2_F | TAAGGGCTGAGCACCAGGAC | qRT-PCR |
| 3,4,5,6 | Ascl2_R | CGTACCAGTCAAGGTGTGCTT | qRT-PCR |

Continued on next page

| Chapter | Oligo | Sequence | Application |
|---|---|---|---|
| 3,4,5,6 | Hand1_F | CAAGATCAAGACTCTGCGCC | qRT-PCR |
| 3,4,5,6 | Hand1_R | TTAATCCTCTTCTCGCCGGG | qRT-PCR |
| 3,4,5,6 | Tpbpa_F | CCAGCACAGCTTTGGACATCA | qRT-PCR |
| 3,4,5,6 | Tpbpa_R | AGCATCCAACTGCGCTTCA | qRT-PCR |
| 3,4,5,6 | Snai1_F | CACGCTGCCTTGTGTCTG | qRT-PCR |
| 3,4,5,6 | Snai1_R | AGACTCTTGGTGCTTGTGGA | qRT-PCR |
| 3,4,5,6 | Gapdh_F | AATGTGTCCGTCGTGGATCT | qRT-PCR |
| 3,4,5,6 | Gapdh_R | CCTGCTTCACCACCTTCTTG | qRT-PCR |
| 4 | Cyr61_F | AGCAGCCTAAAAAAGGGCA | qRT-PCR |
| 4 | Cyr61_R | GGTATTTCTTGACACTGGAGCA | qRT-PCR |
| 4 | Htra1_F | ACGCCAAGACCTACACCAAC | qRT-PCR |
| 4 | Htra1_R | TCTCCACCACATCAGCAATAA | qRT-PCR |
| 4 | Plac1_F | GCACAAAGCCACGTTTCA | qRT-PCR |
| 4 | Plac1_R | GCATGAAGGCGACTAAGACAG | qRT-PCR |
| 4 | Mapk13_F | AAGCGGACAGGGGAGAAG | qRT-PCR |
| 4 | Mapk13_R | AGGGGTGAAGACATCCAGAA | qRT-PCR |
| 4 | Bak1_F | CAACCCCGAGATGGACAA | qRT-PCR |
| 4 | Bak1_R | TGTGGGCTGAAGCTGTTCT | qRT-PCR |
| 4 | Cdh5_F | ATGAGCCCCCTGTCTTCC | qRT-PCR |
| 4 | Cdh5_R | TGGTCTTGCGGATGGAGT | qRT-PCR |
| 4 | Itgb3_F | GTGGAAGAGCCTGAGTGTCC | qRT-PCR |
| 4 | Itgb3_R | TCCAGATGAGCAGAGTAGCAAG | qRT-PCR |
| 4,6 | Cdx2_3'_F | TCCTGCTGACTGCTTTCTGA | qRT-PCR |
| 4,6 | Cdx2_3'_R | CCCTTCCTGATTTGTGGAGA | qRT-PCR |
| 5 | Junb_F | GCAGCTACTTTTCGGGTCAG | qRT-PCR |
| 5 | Junb_R | TTCATCTTGTGCAGGTCGTC | qRT-PCR |
| 5 | Fosb_F | GCAACCCACCCTCATCTCT | qRT-PCR |
| 5 | Fosb_R | GGGCAGACACAGGTCCAC | qRT-PCR |
| 5 | Ppard_F | AGGAGCTGGGGGATTCTG | qRT-PCR |
| 5 | Ppard_R | CCGTGGGTTTGTCTTCATCT | qRT-PCR |
| 5,6 | Syna_F | CCTCACCTCCCAGGCCCCTC | qRT-PCR |
| 5,6 | Syna_R | GGCAGGGAGTTTGCCCACGA | qRT-PCR |
| 5,6 | Prl3d1_F | GGAGCCTACATTGTGGTGGA | qRT-PCR |
| 5,6 | Prl3d1_R | TCCCTATACACATCTGCGGC | qRT-PCR |
| 5,6 | Prl3d3_F | GGCTCCGGAATGCAATTGTT | qRT-PCR |
| 5,6 | Prl3d3_R | TCCCTATACACATCTGCGGC | qRT-PCR |
| 5,6 | Prl2c2_F | TGCTCCTGGATACTGCTCCT | qRT-PCR |
| 5,6 | Prl2c2_R | TGAGACAAACTGCCGGCTAA | qRT-PCR |
| 5,6 | Ctsq_F | AATTGGCTATGGTTATGTGGGA | qRT-PCR |
| 5,6 | Ctsq_R | TCACACAGTAGGGTATTGGG | qRT-PCR |
| 5,6 | Cebpa_F | ACTAACGGCTGGGGGAAG | qRT-PCR |
| 5,6 | Cebpa_R | CTCTGGGAGGCAGACGAA | qRT-PCR |
| 5,6 | Jun_F | ACGACCTTCTACGACGATGC | qRT-PCR |
| 5,6 | Jun_R | CCAGGTTCAAGGTCATGCTC | qRT-PCR |
| 6 | MiSeq_Cdx2ex2_1_F | TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGAAATGACTCCTGGGTTAGGG | MiSeq |
| 6 | MiSeq_Cdx2ex2_1_R | GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCATTGAAAGAGCGGCAAAA | MiSeq |
| 6 | MiSeq_Cdx2ex2_2_F | TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCAAAACTTCTCCTTCCTCCA | MiSeq |
| 6 | MiSeq_Cdx2ex2_2_R | GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCGCTCTTTCTCTGTCCAAGTG | MiSeq |
| 6 | MiSeq_Cdx2ex2_3_F | TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGTGAGTGGATATTGTCAACCTG | MiSeq |
| 6 | MiSeq_Cdx2ex2_3_R | GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACAGCTCCCTCGACCAAAC | MiSeq |

# Appendix C: Enhancer Construct Sequencing Validation

| Enhancer Validation Vectors | Location in mm10 | Size in mm10 | Deviation in sequence from mm10 |
|---|---|---|---|
| pGL3-basic + 209bp Promoter | Chr11:57832121-57831913 | 209bp | Point mutation T→C at position Chr11:57832019 |
| pGL3-enhancer + 209bp Promoter | Chr11:57832121-57831913 | 209bp | Point mutation T→C at position Chr11:57832019 |
| pGL3-basic + 283bp Promoter | Chr11:57832195-57831913 | 283bp | Point mutation T→C at position Chr11:57832019 |
| pGL3-enhancer + 283bp Promoter | Chr11:57832195-57831913 | 283bp | Point mutation T→C at position Chr11:57832019 |
| pGL3-basic + 673bp Promoter | Chr11:57832585-57831913 | 673bp | Point mutation T→C at position Chr11:57832019<br>Loss of 'acacacacac' prior to Chr11:57832397 |
| pGL3-enhancer + 673bp Promoter | Chr11:57832585-57831913 | 673bp | Point mutation T→C at position Chr11:57832019<br>Loss of 'acacacacac' prior to Chr11:57832397 |
| 283bp Hand1 promoter + Region A | Chr11:57850827-57851433 | 607bp | Insertion of 'GAAG' after position Chr11:57850851<br>Point mutation T→C at position Chr11:57851044 |
| 283bp Hand1 promoter + Region B | Chr11:57857626-57857998 | 373bp | Point mutation G→A at position Chr11:57857662<br>Point mutation T→C at position Chr11:57857704<br>Point mutation G→T at position Chr11:57857726<br>Point mutation C→T at position Chr11:57857731<br>Point mutation T→C at position Chr11:57857749<br>Point mutation G→A at position Chr11:57857949<br>Point mutation G→A at position Chr11:57857961 |
| 283bp Hand1 promoter + Region C | Chr11:57864343-57865008 | 666bp | Point mutation C→T at position Chr11:57864754<br>Point mutation G→A at position Chr11:57864769<br>Point mutation G→A at position Chr11:57864774<br>Point mutation A→T at position Chr 11:57864715<br>Gain of 'CTCTCTCTCTCTCTCT' after Chr 11:57864715 in string of repetitive TCs<br>Point mutation T→A at position Chr11:57864508 |
| 283bp Hand1 promoter + Region D | Chr11:57869180-57869527 | 349bp | 1bp deletion at position Chr11:57869202<br>Point mutation T→C at position Chr11:57869352<br>Point mutation G→T at position Chr11:57869478 |
| 283bp Hand1 promoter + Region E | Chr11:57894913-57895360 | 448bp | Point mutation A→G at position Chr11:57894978<br>Point mutation A→Gat position Chr11:57895282<br>Point mutation T→C at position Chr11:57895125<br>Point mutation T→C at position Chr11:57895282<br>Point mutation T→C at position Chr11:57895285<br>Point mutation T→C at position Chr11:57895310<br>Point mutation A→C at position Chr11:57895315<br>Point mutation G→A at position Chr11:57895327 |
| 283bp Hand1 promoter + Region F | Chr11:57896662-57897162 | 501bp | Point mutation T→C at position Chr11:57896849 |
| 283bp Hand1 promoter + Region C1 | Chr11:57881044-57881573 | 538bp | Point mutation A→G at position Chr11:57881130<br>Point mutation T→A at position Chr11:57881284<br>1bp deletion at position Chr11:57881285<br>Point mutation T→C at position Chr11:57881324<br>Point mutation A→G at position Chr11:57881362<br>Point mutation G→A at position Chr11:57881433<br>Point mutation G→T at position Chr11:57881496 |
| 283bp Hand1 promoter + Region C2 | Chr11:57915883-57916496 | 589bp | Point mutation A→T at position Chr11:57915930<br>1bp deletion at Chr11:57916007<br>Point mutation G→A at position Chr11:57916372<br>Point mutation T→C at position Chr11:57916464 |

# Bibliography

ADACHI, K., NIKAIDO, I., OHTA, H., OHTSUKA, S., URA, H., KADOTA, M., WAKAYAMA, T., UEDA, H. R. & NIWA, H. 2013. Context-dependent wiring of Sox2 regulatory networks for self-renewal of embryonic and trophoblast stem cells. *Mol Cell,* 52**,** 380-92.

ADAMSON, S. L., LU, Y., WHITELEY, K. J., HOLMYARD, D., HEMBERGER, M., PFARRER, C. & CROSS, J. C. 2002. Interactions between trophoblast cells and the maternal and fetal circulation in the mouse placenta. *Dev Biol,* 250**,** 358-73.

AIBA, K., NEDOREZOV, T., PIAO, Y., NISHIYAMA, A., MATOBA, R., SHAROVA, L. V., SHAROV, A. A., YAMANAKA, S., NIWA, H. & KO, M. S. 2009. Defining developmental potency and cell lineage trajectories by expression profiling of differentiating mouse embryonic stem cells. *DNA Res,* 16**,** 73-80.

ALARCON, V. B. 2010. Cell polarity regulator PARD6B is essential for trophectoderm formation in the preimplantation mouse embryo. *Biol Reprod,* 83**,** 347-58.

AMITA, M., ADACHI, K., ALEXENKO, A. P., SINHA, S., SCHUST, D. J., SCHULZ, L. C., ROBERTS, R. M. & EZASHI, T. 2013. Complete and unidirectional conversion of human embryonic stem cells to trophoblast by BMP4. *Proc Natl Acad Sci U S A,* 110**,** E1212-21.

ARMAN, E., HAFFNER-KRAUSZ, R., CHEN, Y., HEATH, J. K. & LONAI, P. 1998. Targeted disruption of fibroblast growth factor (FGF) receptor 2 suggests a role for FGF signaling in pregastrulation mammalian development. *Proc Natl Acad Sci U S A,* 95**,** 5082-7.

ARNOLD, S. J., SUGNASEELAN, J., GROSZER, M., SRINIVAS, S. & ROBERTSON, E. J. 2009. Generation and analysis of a mouse line harboring GFP in the Eomes/Tbr2 locus. *Genesis,* 47**,** 775-81.

AUMAN, H. J., NOTTOLI, T., LAKIZA, O., WINGER, Q., DONALDSON, S. & WILLIAMS, T. 2002. Transcription factor AP-2gamma is essential in the extra-embryonic lineages for early postimplantation development. *Development,* 129**,** 2733-47.

AVILION, A. A., NICOLIS, S. K., PEVNY, L. H., PEREZ, L., VIVIAN, N. & LOVELL-BADGE, R. 2003. Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev,* 17**,** 126-40.

BAEK, S., GOLDSTEIN, I. & HAGER, G. L. 2017. Bivariate Genomic Footprinting Detects Changes in Transcription Factor Activity. *Cell Rep,* 19**,** 1710-1722.

BAILEY, T. L., BODEN, M., BUSKE, F. A., FRITH, M., GRANT, C. E., CLEMENTI, L., REN, J., LI, W. W. & NOBLE, W. S. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res,* 37**,** W202-8.

BECK, F., ERLER, T., RUSSELL, A. & JAMES, R. 1995. Expression of Cdx-2 in the mouse embryo and placenta: possible role in patterning of the extra-embryonic membranes. *Dev Dyn,* 204**,** 219-27.

BEDDINGTON, R. S. & ROBERTSON, E. J. 1989. An assessment of the developmental potential of embryonic stem cells in the midgestation mouse embryo. *Development,* 105**,** 733-7.

BELTON, J. M., MCCORD, R. P., GIBCUS, J. H., NAUMOVA, N., ZHAN, Y. & DEKKER, J. 2012. Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods,* 58**,** 268-76.

BENAYOUN, B. A., POLLINA, E. A., UCAR, D., MAHMOUDI, S., KARRA, K., WONG, E. D., DEVARAJAN, K., DAUGHERTY, A. C., KUNDAJE, A. B., MANCINI, E., HITZ, B. C., GUPTA, R., RANDO, T. A., BAKER, J. C., SNYDER, M. P.,

CHERRY, J. M. & BRUNET, A. 2014. H3K4me3 breadth is linked to cell identity and transcriptional consistency. *Cell,* 158**,** 673-88.

BENCHETRIT, H., HERMAN, S., VAN WIETMARSCHEN, N., WU, T., MAKEDONSKI, K., MAOZ, N., YOM TOV, N., STAVE, D., LASRY, R., ZAYAT, V., XIAO, A., LANSDORP, P. M., SEBBAN, S. & BUGANIM, Y. 2015. Extensive Nuclear Reprogramming Underlies Lineage Conversion into Functional Trophoblast Stem-like Cells. *Cell Stem Cell,* 17**,** 543-56.

BERG, D. K., SMITH, C. S., PEARTON, D. J., WELLS, D. N., BROADHURST, R., DONNISON, M. & PFEFFER, P. L. 2011. Trophectoderm lineage determination in cattle. *Dev Cell,* 20**,** 244-55.

BERNARDO, A. S., FAIAL, T., GARDNER, L., NIAKAN, K. K., ORTMANN, D., SENNER, C. E., CALLERY, E. M., TROTTER, M. W., HEMBERGER, M., SMITH, J. C., BARDWELL, L., MOFFETT, A. & PEDERSEN, R. A. 2011. BRACHYURY and CDX2 mediate BMP-induced differentiation of human and mouse pluripotent stem cells into embryonic and extraembryonic lineages. *Cell Stem Cell,* 9**,** 144-55.

BERNSTEIN, B. E., KAMAL, M., LINDBLAD-TOH, K., BEKIRANOV, S., BAILEY, D. K., HUEBERT, D. J., MCMAHON, S., KARLSSON, E. K., KULBOKAS, E. J., 3RD, GINGERAS, T. R., SCHREIBER, S. L. & LANDER, E. S. 2005. Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell,* 120**,** 169-81.

BEVILACQUA, E. M. & ABRAHAMSOHN, P. A. 1988. Ultrastructure of trophoblast giant cell transformation during the invasive stage of implantation of the mouse embryo. *J Morphol,* 198**,** 341-51.

BLAKELEY, P., FOGARTY, N. M., DEL VALLE, I., WAMAITHA, S. E., HU, T. X., ELDER, K., SNELL, P., CHRISTIE, L., ROBSON, P. & NIAKAN, K. K. 2015. Defining the three cell lineages of the human blastocyst by single-cell RNA-seq. *Development,* 142**,** 3151-65.

BLIJ, S., PARENTI, A., TABATABAI-YAZDI, N. & RALSTON, A. 2015. Cdx2 efficiently induces trophoblast stem-like cells in naive, but not primed, pluripotent stem cells. *Stem Cells Dev,* 24**,** 1352-65.

BONEV, B., MENDELSON COHEN, N., SZABO, Q., FRITSCH, L., PAPADOPOULOS, G. L., LUBLING, Y., XU, X., LV, X., HUGNOT, J. P., TANAY, A. & CAVALLI, G. 2017. Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell,* 171**,** 557-572.e24.

BOROVIAK, T., LOOS, R., BERTONE, P., SMITH, A. & NICHOLS, J. 2014. The ability of inner-cell-mass cells to self-renew as embryonic stem cells is acquired following epiblast specification. *Nat Cell Biol,* 16**,** 516-28.

BOU, G., LIU, S., SUN, M., ZHU, J., XUE, B., GUO, J., ZHAO, Y., QU, B., WENG, X., WEI, Y., LEI, L. & LIU, Z. 2017. CDX2 is essential for cell proliferation and polarity in porcine blastocysts. *Development,* 144**,** 1296-1306.

BOYER, L. A., PLATH, K., ZEITLINGER, J., BRAMBRINK, T., MEDEIROS, L. A., LEE, T. I., LEVINE, S. S., WERNIG, M., TAJONAR, A., RAY, M. K., BELL, G. W., OTTE, A. P., VIDAL, M., GIFFORD, D. K., YOUNG, R. A. & JAENISCH, R. 2006. Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature,* 441**,** 349-53.

BRADLEY, A., EVANS, M., KAUFMAN, M. H. & ROBERTSON, E. 1984. Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature,* 309**,** 255-6.

BRAUDE, P., BOLTON, V. & MOORE, S. 1988. Human gene expression first occurs between the four- and eight-cell stages of preimplantation development. *Nature,* 332**,** 459-61.

BUENROSTRO, J. D., GIRESI, P. G., ZABA, L. C., CHANG, H. Y. & GREENLEAF, W. J. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods,* 10**,** 1213-8.

BUENROSTRO, J. D., WU, B., CHANG, H. Y. & GREENLEAF, W. J. 2015. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol,* 109**,** 21 29 1-9.

BURGLIN, T. R. & AFFOLTER, M. 2016. Homeodomain proteins: an update. *Chromosoma,* 125**,** 497-521.

BUSHNELL, B. 2016. BBMap short-read aligner, and other bioinformatics tools. .

CAMBULI, F., MURRAY, A., DEAN, W., DUDZINSKA, D., KRUEGER, F., ANDREWS, S., SENNER, C. E., COOK, S. J. & HEMBERGER, M. 2014. Epigenetic memory of the first cell fate decision prevents complete ES cell reprogramming into trophoblast. *Nat Commun,* 5**,** 5538.

CAO, Z., CAREY, T. S., GANGULY, A., WILSON, C. A., PAUL, S. & KNOTT, J. G. 2015. Transcription factor AP-2gamma induces early Cdx2 expression and represses HIPPO signaling to specify the trophectoderm lineage. *Development,* 142**,** 1606-15.

CARNEY, E. W., PRIDEAUX, V., LYE, S. J. & ROSSANT, J. 1993. Progressive expression of trophoblast-specific genes during formation of mouse trophoblast giant cells in vitro. *Mol Reprod Dev,* 34**,** 357-68.

CARR, A. & BIGGIN, M. D. 1999. A comparison of in vivo and in vitro DNA-binding specificities suggests a new model for homeoprotein DNA binding in Drosophila embryos. *Embo j,* 18**,** 1598-608.

CARTEGNI, L., CHEW, S. L. & KRAINER, A. R. 2002. Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet,* 3**,** 285-98.

CARTEGNI, L., WANG, J., ZHU, Z., ZHANG, M. Q. & KRAINER, A. R. 2003. ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res,* 31**,** 3568-71.

CATRON, K. M., ILER, N. & ABATE, C. 1993. Nucleotides flanking a conserved TAAT core dictate the DNA binding specificity of three murine homeodomain proteins. *Mol Cell Biol,* 13**,** 2354-65.

CHAWENGSAKSOPHAK, K., DE GRAAFF, W., ROSSANT, J., DESCHAMPS, J. & BECK, F. 2004. Cdx2 is essential for axial elongation in mouse development. *Proc Natl Acad Sci U S A,* 101**,** 7641-5.

CHAWENGSAKSOPHAK, K., JAMES, R., HAMMOND, V. E., KONTGEN, F. & BECK, F. 1997. Homeosis and intestinal tumours in Cdx2 mutant mice. *Nature,* 386**,** 84-7.

CHAZAUD, C. & YAMANAKA, Y. 2016. Lineage specification in the mouse preimplantation embryo. *Development,* 143**,** 1063-74.

CHEN, A. E., EGLI, D., NIAKAN, K., DENG, J., AKUTSU, H., YAMAKI, M., COWAN, C., FITZ-GERALD, C., ZHANG, K., MELTON, D. A. & EGGAN, K. 2009a. Optimal timing of inner cell mass isolation increases the efficiency of human embryonic stem cell derivation and allows generation of sibling cell lines. *Cell Stem Cell,* 4**,** 103-6.

CHEN, L., YABUUCHI, A., EMINLI, S., TAKEUCHI, A., LU, C. W., HOCHEDLINGER, K. & DALEY, G. Q. 2009b. Cross-regulation of the Nanog and Cdx2 promoters. *Cell Res,* 19**,** 1052-61.

CHOI, H. J., SANDERS, T. A., TORMOS, K. V., AMERI, K., TSAI, J. D., PARK, A. M., GONZALEZ, J., RAJAH, A. M., LIU, X., QUINONEZ, D. M., RINAUDO, P. F. & MALTEPE, E. 2013. ECM-dependent HIF induction directs trophoblast stem

cell fate via LIMK1-mediated cytoskeletal rearrangement. *PLoS One,* 8**,** e56949.
CHOI, I., CAREY, T. S., WILSON, C. A. & KNOTT, J. G. 2012. Transcription factor AP-2gamma is a core regulator of tight junction biogenesis and cavity formation during mouse early embryogenesis. *Development,* 139**,** 4623-32.
CHUONG, E. B., RUMI, M. A., SOARES, M. J. & BAKER, J. C. 2013. Endogenous retroviruses function as species-specific enhancer elements in the placenta. *Nat Genet,* 45**,** 325-9.
CIRILLO, L. A., LIN, F. R., CUESTA, I., FRIEDMAN, D., JARNIK, M. & ZARET, K. S. 2002. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol Cell,* 9**,** 279-89.
CIRUNA, B. G. & ROSSANT, J. 1999. Expression of the T-box gene Eomesodermin during early mouse development. *Mech Dev,* 81**,** 199-203.
COAN, P. M., CONROY, N., BURTON, G. J. & FERGUSON-SMITH, A. C. 2006. Origin and characteristics of glycogen cells in the developing murine placenta. *Dev Dyn,* 235**,** 3280-94.
COAN, P. M., FERGUSON-SMITH, A. C. & BURTON, G. J. 2005. Ultrastructural changes in the interhaemal membrane and junctional zone of the murine chorioallantoic placenta across gestation. *J Anat,* 207**,** 783-96.
COCKBURN, K., BIECHELE, S., GARNER, J. & ROSSANT, J. 2013. The Hippo pathway member Nf2 is required for inner cell mass specification. *Curr Biol,* 23**,** 1195-201.
CONG, L., RAN, F. A., COX, D., LIN, S., BARRETTO, R., HABIB, N., HSU, P. D., WU, X., JIANG, W., MARRAFFINI, L. A. & ZHANG, F. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science,* 339**,** 819-23.
CONWAY, J. R., LEX, A. & GEHLENBORG, N. 2017. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics,* 33**,** 2938-2940.
CRAWFORD, G. E., HOLT, I. E., WHITTLE, J., WEBB, B. D., TAI, D., DAVIS, S., MARGULIES, E. H., CHEN, Y., BERNAT, J. A., GINSBURG, D., ZHOU, D., LUO, S., VASICEK, T. J., DALY, M. J., WOLFSBERG, T. G. & COLLINS, F. S. 2006. Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS). *Genome Res,* 16**,** 123-31.
CREMER, T. & CREMER, M. 2010. Chromosome territories. *Cold Spring Harb Perspect Biol,* 2**,** a003889.
CREYGHTON, M. P., CHENG, A. W., WELSTEAD, G. G., KOOISTRA, T., CAREY, B. W., STEINE, E. J., HANNA, J., LODATO, M. A., FRAMPTON, G. M., SHARP, P. A., BOYER, L. A., YOUNG, R. A. & JAENISCH, R. 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A,* 107**,** 21931-6.
CROSS, J. C. 2014. More of a good thing or less of a bad thing: gene copy number variation in polyploid cells of the placenta. *PLoS Genet,* 10**,** e1004330.
CROSS, J. C., FLANNERY, M. L., BLANAR, M. A., STEINGRIMSSON, E., JENKINS, N. A., COPELAND, N. G., RUTTER, W. J. & WERB, Z. 1995. Hxt encodes a basic helix-loop-helix transcription factor that regulates trophoblast cell development. *Development,* 121**,** 2513-23.
CROSS, J. C., NAKANO, H., NATALE, D. R., SIMMONS, D. G. & WATSON, E. D. 2006. Branching morphogenesis during development of placental villi. *Differentiation,* 74**,** 393-401.
CROSS, J. C., WERB, Z. & FISHER, S. J. 1994. Implantation and the placenta: key pieces of the development puzzle. *Science,* 266**,** 1508-18.
DAS, P., EZASHI, T., SCHULZ, L. C., WESTFALL, S. D., LIVINGSTON, K. A. & ROBERTS, R. M. 2007. Effects of fgf2 and oxygen in the bmp4-driven

differentiation of trophoblast from human embryonic stem cells. *Stem Cell Res,* 1**,** 61-74.
DAUDE, N., WOHLGEMUTH, S., BROWN, R., PITSTICK, R., GAPESHINA, H., YANG, J., CARLSON, G. A. & WESTAWAY, D. 2012. Knockout of the prion protein (PrP)-like Sprn gene does not produce embryonic lethality in combination with PrP(C)-deficiency. *Proc Natl Acad Sci U S A,* 109**,** 9035-40.
DE PAEPE, C., CAUFFMAN, G., VERLOES, A., STERCKX, J., DEVROEY, P., TOURNAYE, H., LIEBAERS, I. & VAN DE VELDE, H. 2013. Human trophectoderm cells are not yet committed. *Hum Reprod,* 28**,** 740-9.
DE SOUZA, A. T., DAI, X., SPENCER, A. G., REPPEN, T., MENZIE, A., ROESCH, P. L., HE, Y., CAGUYONG, M. J., BLOOMER, S., HERWEIJER, H., WOLFF, J. A., HAGSTROM, J. E., LEWIS, D. L., LINSLEY, P. S. & ULRICH, R. G. 2006. Transcriptional and phenotypic comparisons of Ppara knockout and siRNA knockdown mice. *Nucleic Acids Res,* 34**,** 4486-94.
DE VRIES, W. N., EVSIKOV, A. V., HAAC, B. E., FANCHER, K. S., HOLBROOK, A. E., KEMLER, R., SOLTER, D. & KNOWLES, B. B. 2004. Maternal beta-catenin and E-cadherin in mouse development. *Development,* 131**,** 4435-45.
DICKSON, A. D. 1966. The form of the mouse blastocyst. *J Anat,* 100**,** 335-48.
DOBIN, A., DAVIS, C. A., SCHLESINGER, F., DRENKOW, J., ZALESKI, C., JHA, S., BATUT, P., CHAISSON, M. & GINGERAS, T. R. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics,* 29**,** 15-21.
DONNISON, M., BEATON, A., DAVEY, H. W., BROADHURST, R., L'HUILLIER, P. & PFEFFER, P. L. 2005. Loss of the extraembryonic ectoderm in Elf5 mutants leads to defects in embryonic patterning. *Development,* 132**,** 2299-308.
DONNISON, M., BROADHURST, R. & PFEFFER, P. L. 2015. Elf5 and Ets2 maintain the mouse extraembryonic ectoderm in a dosage dependent synergistic manner. *Dev Biol,* 397**,** 77-88.
DORIGHI, K. M., SWIGUT, T., HENRIQUES, T., BHANU, N. V., SCRUGGS, B. S., NADY, N., STILL, C. D., 2ND, GARCIA, B. A., ADELMAN, K. & WYSOCKA, J. 2017. Mll3 and Mll4 Facilitate Enhancer RNA Synthesis and Transcription from Promoters Independently of H3K4 Monomethylation. *Mol Cell,* 66**,** 568-576.e4.
DOUDNA, J. A. & CHARPENTIER, E. 2014. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science,* 346**,** 1258096.
DRAGAN, A. I., LI, Z., MAKEYEVA, E. N., MILGOTINA, E. I., LIU, Y., CRANE-ROBINSON, C. & PRIVALOV, P. L. 2006. Forces driving the binding of homeodomains to DNA. *Biochemistry,* 45**,** 141-51.
DUPRESSOIR, A., MARCEAU, G., VERNOCHET, C., BENIT, L., KANELLOPOULOS, C., SAPIN, V. & HEIDMANN, T. 2005. Syncytin-A and syncytin-B, two fusogenic placenta-specific murine envelope genes of retroviral origin conserved in Muridae. *Proc Natl Acad Sci U S A,* 102**,** 725-30.
DUPREY, P., CHOWDHURY, K., DRESSLER, G. R., BALLING, R., SIMON, D., GUENET, J. L. & GRUSS, P. 1988. A mouse gene homologous to the Drosophila gene caudal is expressed in epithelial cells from the embryonic intestine. *Genes Dev,* 2**,** 1647-54.
EDELSTEIN, A. D., TSUCHIDA, M. A., AMODAJ, N., PINKARD, H., VALE, R. D. & STUURMAN, N. 2014. Advanced methods of microscope control using muManager software. *J Biol Methods,* 1.
EDWARDS, R. G., PURDY, J. M., STEPTOE, P. C. & WALTERS, D. E. 1981. The growth of human preimplantation embryos in vitro. *Am J Obstet Gynecol,* 141**,** 408-16.
EL-BROLOSY, M. A., KONTARAKIS, Z., ROSSI, A., KUENNE, C., GUNTHER, S., FUKUDA, N., KIKHI, K., BOEZIO, G. L. M., TAKACS, C. M., LAI, S. L., FUKUDA, R., GERRI, C., GIRALDEZ, A. J. & STAINIER, D. Y. R. 2019.

Genetic compensation triggered by mutant mRNA degradation. *Nature,* 568**,** 193-197.

ENDERS, A. C. 1965. A COMPARATIVE STUDY OF THE FINE STRUCTURE OF THE TROPHOBLAST IN SEVERAL HEMOCHORIAL PLACENTAS. *Am J Anat,* 116**,** 29-67.

ERLEBACHER, A., PRICE, K. A. & GLIMCHER, L. H. 2004. Maintenance of mouse trophoblast stem cell proliferation by TGF-beta/activin. *Dev Biol,* 275**,** 158-69.

EVANS, M. J., BRADLEY, A., KUEHN, M. R. & ROBERTSON, E. J. 1985. The ability of EK cells to form chimeras after selection of clones in G418 and some observations on the integration of retroviral vector proviral DNA into EK cells. *Cold Spring Harb Symp Quant Biol,* 50**,** 685-9.

EVANS, M. J. & KAUFMAN, M. H. 1981. Establishment in culture of pluripotential cells from mouse embryos. *Nature,* 292**,** 154-6.

EZHKOVA, E., PASOLLI, H. A., PARKER, J. S., STOKES, N., SU, I. H., HANNON, G., TARAKHOVSKY, A. & FUCHS, E. 2009. Ezh2 orchestrates gene expression for the stepwise differentiation of tissue-specific stem cells. *Cell,* 136**,** 1122-35.

FELDMAN, B., POUEYMIROU, W., PAPAIOANNOU, V. E., DECHIARA, T. M. & GOLDFARB, M. 1995. Requirement of FGF-4 for postimplantation mouse development. *Science,* 267**,** 246-9.

FERNER, K. & MESS, A. 2011. Evolution and development of fetal membranes and placentation in amniote vertebrates. *Respir Physiol Neurobiol,* 178**,** 39-50.

FERRIER, D. E., DEWAR, K., COOK, A., CHANG, J. L., HILL-FORCE, A. & AMEMIYA, C. 2005. The chordate ParaHox cluster. *Curr Biol,* 15**,** R820-2.

FESTUCCIA, N., OSORNO, R., HALBRITTER, F., KARWACKI-NEISIUS, V., NAVARRO, P., COLBY, D., WONG, F., YATES, A., TOMLINSON, S. R. & CHAMBERS, I. 2012. Esrrb is a direct Nanog target gene that can substitute for Nanog function in pluripotent cells. *Cell Stem Cell,* 11**,** 477-90.

FIERRO-GONZALEZ, J. C., WHITE, M. D., SILVA, J. C. & PLACHTA, N. 2013. Cadherin-dependent filopodia control preimplantation embryo compaction. *Nat Cell Biol,* 15**,** 1424-33.

FLACH, G., JOHNSON, M. H., BRAUDE, P. R., TAYLOR, R. A. & BOLTON, V. N. 1982. The transition from maternal to embryonic control in the 2-cell mouse embryo. *Embo j,* 1**,** 681-6.

FRIAS-ALDEGUER, J., KIP, M., VIVIÉ, J., LI, L., ALEMANY, A., KORVING, J., DARMIS, F., VAN OUDENAARDEN, A., VAN BLITTERSWIJK, C. A., GEIJSEN, N. & RIVRON, N. C. 2019. Embryonic signals perpetuate polar-like trophoblast stem cells and pattern the blastocyst axis. *bioRxiv***,** 510362.

FRUM, T., MURPHY, T. M. & RALSTON, A. 2018. HIPPO signaling resolves embryonic cell fate conflicts during establishment of pluripotency in vivo. *Elife,* 7.

FUJIMORI, T., KUROTAKI, Y., MIYAZAKI, J. & NABESHIMA, Y. 2003. Analysis of cell lineage in two- and four-cell mouse embryos. *Development,* 130**,** 5113-22.

GAGNON, R. 2003. Placental insufficiency and its consequences. *Eur J Obstet Gynecol Reprod Biol,* 110 Suppl 1**,** S99-107.

GALAS, D. J. & SCHMITZ, A. 1978. DNAse footprinting: a simple method for the detection of protein-DNA binding specificity. *Nucleic Acids Res,* 5**,** 3157-70.

GAMER, L. W. & WRIGHT, C. V. 1993. Murine Cdx-4 bears striking similarities to the Drosophila caudal gene in its homeodomain sequence and early expression pattern. *Mech Dev,* 43**,** 71-81.

GAO, H., GAO, R., ZHANG, L., XIU, W., ZANG, R., WANG, H., ZHANG, Y., CHEN, J., GAO, Y. & GAO, S. 2019. Esrrb plays important roles in maintaining self-renewal of trophoblast stem cells (TSCs) and reprogramming somatic cells to induced TSCs. *J Mol Cell Biol,* 11**,** 463-473.

GAO, N., WHITE, P. & KAESTNER, K. H. 2009. Establishment of intestinal identity and epithelial-mesenchymal signaling by Cdx2. *Dev Cell,* 16**,** 588-99.

GARDNER, R. L., PAPAIOANNOU, V. E. & BARTON, S. C. 1973. Origin of the ectoplacental cone and secondary giant cells in mouse blastocysts reconstituted from isolated trophoblast and inner cell mass. *J Embryol Exp Morphol,* 30**,** 561-72.

GENG, Y., YU, Q., SICINSKA, E., DAS, M., SCHNEIDER, J. E., BHATTACHARYA, S., RIDEOUT, W. M., BRONSON, R. T., GARDNER, H. & SICINSKI, P. 2003. Cyclin E ablation in the mouse. *Cell,* 114**,** 431-43.

GENTSCH, G. E. & SMITH, J. C. 2017. Efficient Preparation of High-Complexity ChIP-Seq Profiles from Early Xenopus Embryos. *Methods Mol Biol,* 1507**,** 23-42.

GENTSCH, G. E., SPRUCE, T., MONTEIRO, R. S., OWENS, N. D. L., MARTIN, S. R. & SMITH, J. C. 2018. Innate Immune Response and Off-Target Mis-splicing Are Common Morpholino-Induced Side Effects in Xenopus. *Dev Cell,* 44**,** 597-610.e10.

GEORGE, K. M., LEONARD, M. W., ROTH, M. E., LIEUW, K. H., KIOUSSIS, D., GROSVELD, F. & ENGEL, J. D. 1994. Embryonic expression and cloning of the murine GATA-3 gene. *Development,* 120**,** 2673-86.

GEORGIADES, P. & ROSSANT, J. 2006. Ets2 is necessary in trophoblast for normal embryonic anteroposterior axis development. *Development,* 133**,** 1059-68.

GIRESI, P. G., KIM, J., MCDANIELL, R. M., IYER, V. R. & LIEB, J. D. 2007. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res,* 17**,** 877-85.

GOOKIN, S., MIN, M., PHADKE, H., CHUNG, M., MOSER, J., MILLER, I., CARTER, D. & SPENCER, S. L. 2017. A map of protein dynamics during cell-cycle progression and cell-cycle exit. *PLoS Biol,* 15**,** e2003268.

GOUTI, M., DELILE, J., STAMATAKI, D., WYMEERSCH, F. J., HUANG, Y., KLEINJUNG, J., WILSON, V. & BRISCOE, J. 2017. A Gene Regulatory Network Balances Neural and Mesoderm Specification during Vertebrate Trunk Development. *Dev Cell,* 41**,** 243-261 e7.

GRAVELEY, B. R. 2000. Sorting out the complexity of SR protein functions. *Rna,* 6**,** 1197-211.

GRONTVED, L., WATERFALL, J. J., KIM, D. W., BAEK, S., SUNG, M. H., ZHAO, L., PARK, J. W., NIELSEN, R., WALKER, R. L., ZHU, Y. J., MELTZER, P. S., HAGER, G. L. & CHENG, S. Y. 2015. Transcriptional activation by the thyroid hormone receptor through ligand-dependent receptor recruitment and chromatin remodelling. *Nat Commun,* 6**,** 7048.

GUILLEMOT, F., NAGY, A., AUERBACH, A., ROSSANT, J. & JOYNER, A. L. 1994. Essential role of Mash-2 in extraembryonic development. *Nature,* 371**,** 333-6.

GUO, G., HUSS, M., TONG, G. Q., WANG, C., LI SUN, L., CLARKE, N. D. & ROBSON, P. 2010. Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. *Dev Cell,* 18**,** 675-85.

GUZMAN-AYALA, M., BEN-HAIM, N., BECK, S. & CONSTAM, D. B. 2004. Nodal protein processing and fibroblast growth factor 4 synergize to maintain a trophoblast stem cell microenvironment. *Proc Natl Acad Sci U S A,* 101**,** 15656-60.

HAMATANI, T., CARTER, M. G., SHAROV, A. A. & KO, M. S. 2004. Dynamics of global gene expression changes during mouse preimplantation development. *Dev Cell,* 6**,** 117-31.

HANCOCK, S. N., AGULNIK, S. I., SILVER, L. M. & PAPAIOANNOU, V. E. 1999. Mapping and expression analysis of the mouse ortholog of Xenopus Eomesodermin. *Mech Dev,* 81**,** 205-8.

HANNIBAL, R. L. & BAKER, J. C. 2016. Selective Amplification of the Genome Surrounding Key Placental Genes in Trophoblast Giant Cells. *Curr Biol,* 26**,** 230-236.

HANNIBAL, R. L., CHUONG, E. B., RIVERA-MULIA, J. C., GILBERT, D. M., VALOUEV, A. & BAKER, J. C. 2014. Copy number variation is a fundamental aspect of the placental genome. *PLoS Genet,* 10**,** e1004290.

HAYAKAWA, K., HIMENO, E., TANAKA, S. & KUNATH, T. 2015. Isolation and manipulation of mouse trophoblast stem cells. *Curr Protoc Stem Cell Biol,* 32**,** 1E 4 1-32.

HAYASHI, Y., FURUE, M. K., TANAKA, S., HIROSE, M., WAKISAKA, N., DANNO, H., OHNUMA, K., OEDA, S., AIHARA, Y., SHIOTA, K., OGURA, A., ISHIURA, S. & ASASHIMA, M. 2010. BMP4 induction of trophoblast from mouse embryonic stem cells in defined culture conditions on laminin. *In Vitro Cell Dev Biol Anim,* 46**,** 416-30.

HE, H. H., MEYER, C. A., HU, S. S., CHEN, M. W., ZANG, C., LIU, Y., RAO, P. K., FEI, T., XU, H., LONG, H., LIU, X. S. & BROWN, M. 2014. Refined DNase-seq protocol and data analysis reveals intrinsic bias in transcription factor footprint identification. *Nat Methods,* 11**,** 73-78.

HE, S., PANT, D., SCHIFFMACHER, A., MEECE, A. & KEEFER, C. L. 2008. Lymphoid enhancer factor 1-mediated Wnt signaling promotes the initiation of trophoblast lineage differentiation in mouse embryonic stem cells. *Stem Cells,* 26**,** 842-9.

HEINTZMAN, N. D., STUART, R. K., HON, G., FU, Y., CHING, C. W., HAWKINS, R. D., BARRERA, L. O., VAN CALCAR, S., QU, C., CHING, K. A., WANG, W., WENG, Z., GREEN, R. D., CRAWFORD, G. E. & REN, B. 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet,* 39**,** 311-8.

HEINZ, S., BENNER, C., SPANN, N., BERTOLINO, E., LIN, Y. C., LASLO, P., CHENG, J. X., MURRE, C., SINGH, H. & GLASS, C. K. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell,* 38**,** 576-89.

HEMBERGER, M. 2008. IFPA award in placentology lecture - characteristics and significance of trophoblast giant cells. *Placenta,* 29 Suppl A**,** S4-9.

HEMBERGER, M., UDAYASHANKAR, R., TESAR, P., MOORE, H. & BURTON, G. J. 2010. ELF5-enforced transcriptional networks define an epigenetically regulated trophoblast stem cell compartment in the human placenta. *Hum Mol Genet,* 19**,** 2456-67.

HERNANDEZ-VERDUN, D. 1974. Morphogenesis of the syncytium in the mouse placenta. Ultrastructural study. *Cell Tissue Res,* 148**,** 381-96.

HIRATE, Y., HIRAHARA, S., INOUE, K., SUZUKI, A., ALARCON, V. B., AKIMOTO, K., HIRAI, T., HARA, T., ADACHI, M., CHIDA, K., OHNO, S., MARIKAWA, Y., NAKAO, K., SHIMONO, A. & SASAKI, H. 2013. Polarity-dependent distribution of angiomotin localizes Hippo signaling in preimplantation embryos. *Curr Biol,* 23**,** 1181-94.

HOME, P., KUMAR, R. P., GANGULY, A., SAHA, B., MILANO-FOSTER, J., BHATTACHARYA, B., RAY, S., GUNEWARDENA, S., PAUL, A., CAMPER, S. A., FIELDS, P. E. & PAUL, S. 2017. Genetic redundancy of GATA factors in the extraembryonic trophoblast lineage ensures the progression of preimplantation and postimplantation mammalian development. *Development,* 144**,** 876-888.

HOME, P., RAY, S., DUTTA, D., BRONSHTEYN, I., LARSON, M. & PAUL, S. 2009. GATA3 is selectively expressed in the trophectoderm of peri-implantation embryo and directly regulates Cdx2 gene expression. *J Biol Chem,* 284**,** 28729-37.

HOME, P., SAHA, B., RAY, S., DUTTA, D., GUNEWARDENA, S., YOO, B., PAL, A., VIVIAN, J. L., LARSON, M., PETROFF, M., GALLAGHER, P. G., SCHULZ, V. P., WHITE, K. L., GOLOS, T. G., BEHR, B. & PAUL, S. 2012. Altered subcellular localization of transcription factor TEAD4 regulates first mammalian cell lineage commitment. *Proc Natl Acad Sci U S A,* 109**,** 7362-7.

HU, D. & CROSS, J. C. 2010. Development and function of trophoblast giant cells in the rodent placenta. *Int J Dev Biol,* 54**,** 341-54.

HUANG, D., GUO, G., YUAN, P., RALSTON, A., SUN, L., HUSS, M., MISTRI, T., PINELLO, L., NG, H. H., YUAN, G., JI, J., ROSSANT, J., ROBSON, P. & HAN, X. 2017. The role of Cdx2 as a lineage specific transcriptional repressor for pluripotent network during the first developmental cell lineage segregation. *Sci Rep,* 7**,** 17156.

HUGHES, M., DOBRIC, N., SCOTT, I. C., SU, L., STAROVIC, M., ST-PIERRE, B., EGAN, S. E., KINGDOM, J. C. & CROSS, J. C. 2004. The Hand1, Stra13 and Gcm1 transcription factors override FGF signaling to promote terminal differentiation of trophoblast stem cells. *Dev Biol,* 271**,** 26-37.

INOUE, F., KIRCHER, M., MARTIN, B., COOPER, G. M., WITTEN, D. M., MCMANUS, M. T., AHITUV, N. & SHENDURE, J. 2017. A systematic comparison reveals substantial differences in chromosomal versus episomal encoding of enhancer activity. *Genome Res,* 27**,** 38-52.

IVANOVA, N., DOBRIN, R., LU, R., KOTENKO, I., LEVORSE, J., DECOSTE, C., SCHAFER, X., LUN, Y. & LEMISCHKA, I. R. 2006. Dissecting self-renewal in stem cells with RNA interference. *Nature,* 442**,** 533-8.

IYER, V. R., HORAK, C. E., SCAFE, C. S., BOTSTEIN, D., SNYDER, M. & BROWN, P. O. 2001. Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature,* 409**,** 533-8.

JACKMAN, S. M., KONG, X. & FANT, M. E. 2012. Plac1 (placenta-specific 1) is essential for normal placental and embryonic development. *Mol Reprod Dev,* 79**,** 564-72.

JAMES, R. & KAZENWADEL, J. 1991. Homeobox gene expression in the intestinal epithelium of adult mice. *J Biol Chem,* 266**,** 3246-51.

JOLMA, A., YIN, Y., NITTA, K. R., DAVE, K., POPOV, A., TAIPALE, M., ENGE, M., KIVIOJA, T., MORGUNOVA, E. & TAIPALE, J. 2015. DNA-dependent formation of transcription factor pairs alters their binding specificity. *Nature,* 527**,** 384-8.

JOSEPH, R., ORLOV, Y. L., HUSS, M., SUN, W., KONG, S. L., UKIL, L., PAN, Y. F., LI, G., LIM, M., THOMSEN, J. S., RUAN, Y., CLARKE, N. D., PRABHAKAR, S., CHEUNG, E. & LIU, E. T. 2010. Integrative model of genomic factors for determining binding site selection by estrogen receptor-alpha. *Mol Syst Biol,* 6**,** 456.

KAFRI, R., SPRINGER, M. & PILPEL, Y. 2009. Genetic redundancy: new tricks for old genes. *Cell,* 136**,** 389-92.

KAISER, S., KOCH, Y., KUHNEL, E., SHARMA, N., GELLHAUS, A., KUCKENBERG, P., SCHORLE, H. & WINTERHAGER, E. 2015. Reduced Gene Dosage of Tfap2c Impairs Trophoblast Lineage Differentiation and Alters Maternal Blood Spaces in the Mouse Placenta. *Biol Reprod,* 93**,** 31.

KANEKO, K. J. & DEPAMPHILIS, M. L. 2013. TEAD4 establishes the energy homeostasis essential for blastocoel formation. *Development,* 140**,** 3680-90.

KAPAHNKE, M., BANNING, A. & TIKKANEN, R. 2016. Random Splicing of Several Exons Caused by a Single Base Change in the Target Exon of CRISPR/Cas9 Mediated Gene Knockout. *Cells,* 5.

KAROLCHIK, D., HINRICHS, A. S., FUREY, T. S., ROSKIN, K. M., SUGNET, C. W., HAUSSLER, D. & KENT, W. J. 2004. The UCSC Table Browser data retrieval tool. *Nucleic Acids Res,* 32**,** D493-6.

KELLY, S. J. 1977. Studies of the developmental potential of 4- and 8-cell stage mouse blastomeres. *J Exp Zool,* 200**,** 365-76.

KENT, W. J., ZWEIG, A. S., BARBER, G., HINRICHS, A. S. & KAROLCHIK, D. 2010. BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics,* 26**,** 2204-7.

KERAMARI, M., RAZAVI, J., INGMAN, K. A., PATSCH, C., EDENHOFER, F., WARD, C. M. & KIMBER, S. J. 2010. Sox2 is essential for formation of trophectoderm in the preimplantation embryo. *PLoS One,* 5**,** e13952.

KHAN, A., FORNES, O., STIGLIANI, A., GHEORGHE, M., CASTRO-MONDRAGON, J. A., VAN DER LEE, R., BESSY, A., CHENEBY, J., KULKARNI, S. R., TAN, G., BARANASIC, D., ARENILLAS, D. J., SANDELIN, A., VANDEPOELE, K., LENHARD, B., BALLESTER, B., WASSERMAN, W. W., PARCY, F. & MATHELIER, A. 2018. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res,* 46**,** D260-d266.

KIBSCHULL, M., NASSIRY, M., DUNK, C., GELLHAUS, A., QUINN, J. A., ROSSANT, J., LYE, S. J. & WINTERHAGER, E. 2004. Connexin31-deficient trophoblast stem cells: a model to analyze the role of gap junction communication in mouse placental development. *Dev Biol,* 273**,** 63-75.

KIBSCHULL, M. & WINTERHAGER, E. 2006. Connexins and trophoblast cell lineage development. *Methods Mol Med,* 121**,** 149-58.

KIDDER, B. L. & PALMER, S. 2010. Examination of transcriptional networks reveals an important role for TCFAP2C, SMARCA4, and EOMES in trophoblast stem cell maintenance. *Genome Res,* 20**,** 458-72.

KING, H. W. & KLOSE, R. J. 2017. The pioneer factor OCT4 requires the chromatin remodeller BRG1 to support gene regulatory element function in mouse embryonic stem cells. *Elife,* 6.

KISSINGER, C. R., LIU, B. S., MARTIN-BLANCO, E., KORNBERG, T. B. & PABO, C. O. 1990. Crystal structure of an engrailed homeodomain-DNA complex at 2.8 A resolution: a framework for understanding homeodomain-DNA interactions. *Cell,* 63**,** 579-90.

KOBOLAK, J., KISS, K., POLGAR, Z., MAMO, S., ROGEL-GAILLARD, C., TANCOS, Z., BOCK, I., BAJI, A. G., TAR, K., PIRITY, M. K. & DINNYES, A. 2009. Promoter analysis of the rabbit POU5F1 gene and its expression in preimplantation stage embryos. *BMC Mol Biol,* 10**,** 88.

KUBACZKA, C., SENNER, C., ARAUZO-BRAVO, M. J., SHARMA, N., KUCKENBERG, P., BECKER, A., ZIMMER, A., BRUSTLE, O., PEITZ, M., HEMBERGER, M. & SCHORLE, H. 2014. Derivation and maintenance of murine trophoblast stem cells under defined conditions. *Stem Cell Reports,* 2**,** 232-42.

KUBACZKA, C., SENNER, C. E., CIERLITZA, M., ARAUZO-BRAVO, M. J., KUCKENBERG, P., PEITZ, M., HEMBERGER, M. & SCHORLE, H. 2015. Direct Induction of Trophoblast Stem Cells from Murine Fibroblasts. *Cell Stem Cell,* 17**,** 557-68.

KUBOTA, K., KENT, L. N., RUMI, M. A., ROBY, K. F. & SOARES, M. J. 2015. Dynamic Regulation of AP-1 Transcriptional Complexes Directs Trophoblast Differentiation. *Mol Cell Biol,* 35**,** 3163-77.

KUCKENBERG, P., BUHL, S., WOYNECKI, T., VAN FURDEN, B., TOLKUNOVA, E., SEIFFE, F., MOSER, M., TOMILIN, A., WINTERHAGER, E. & SCHORLE, H. 2010. The transcription factor TCFAP2C/AP-2gamma cooperates with CDX2 to maintain trophectoderm formation. *Mol Cell Biol,* 30**,** 3310-20.

KUCKENBERG, P., KUBACZKA, C. & SCHORLE, H. 2012. The role of transcription factor Tcfap2c/TFAP2C in trophectoderm development. *Reprod Biomed Online,* 25**,** 12-20.

KUCKENBERG, P., PEITZ, M., KUBACZKA, C., BECKER, A., EGERT, A., WARDELMANN, E., ZIMMER, A., BRUSTLE, O. & SCHORLE, H. 2011. Lineage conversion of murine extraembryonic trophoblast stem cells to pluripotent stem cells. *Mol Cell Biol,* 31**,** 1748-56.

KUIJK, E. W., DU PUY, L., VAN TOL, H. T., OEI, C. H., HAAGSMAN, H. P., COLENBRANDER, B. & ROELEN, B. A. 2008. Differences in early lineage segregation between mammals. *Dev Dyn,* 237**,** 918-27.

KUMAR, N., TSAI, Y. H., CHEN, L., ZHOU, A., BANERJEE, K. K., SAXENA, M., HUANG, S., TOKE, N. H., XING, J., SHIVDASANI, R. A., SPENCE, J. R. & VERZI, M. P. 2019. The lineage-specific transcription factor CDX2 navigates dynamic chromatin to control distinct stages of intestine development. *Development,* 146.

KUMAR, R. P., RAY, S., HOME, P., SAHA, B., BHATTACHARYA, B., WILKINS, H. M., CHAVAN, H., GANGULY, A., MILANO-FOSTER, J., PAUL, A., KRISHNAMURTHY, P., SWERDLOW, R. H. & PAUL, S. 2018. Regulation of energy metabolism during early mammalian development: TEAD4 controls mitochondrial transcription. *Development,* 145.

KUMAR, S. & BUCHER, P. 2016. Predicting transcription factor site occupancy using DNA sequence intrinsic and cell-type specific chromatin features. *BMC Bioinformatics,* 17 Suppl 1**,** 4.

KUNATH, T., YAMANAKA, Y., DETMAR, J., MACPHEE, D., CANIGGIA, I., ROSSANT, J. & JURISICOVA, A. 2014. Developmental differences in the expression of FGF receptors between human and mouse embryos. *Placenta,* 35**,** 1079-88.

LANDT, S. G., MARINOV, G. K., KUNDAJE, A., KHERADPOUR, P., PAULI, F., BATZOGLOU, S., BERNSTEIN, B. E., BICKEL, P., BROWN, J. B., CAYTING, P., CHEN, Y., DESALVO, G., EPSTEIN, C., FISHER-AYLOR, K. I., EUSKIRCHEN, G., GERSTEIN, M., GERTZ, J., HARTEMINK, A. J., HOFFMAN, M. M., IYER, V. R., JUNG, Y. L., KARMAKAR, S., KELLIS, M., KHARCHENKO, P. V., LI, Q., LIU, T., LIU, X. S., MA, L., MILOSAVLJEVIC, A., MYERS, R. M., PARK, P. J., PAZIN, M. J., PERRY, M. D., RAHA, D., REDDY, T. E., ROZOWSKY, J., SHORESH, N., SIDOW, A., SLATTERY, M., STAMATOYANNOPOULOS, J. A., TOLSTORUKOV, M. Y., WHITE, K. P., XI, S., FARNHAM, P. J., LIEB, J. D., WOLD, B. J. & SNYDER, M. 2012. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res,* 22**,** 1813-31.

LATOS, P. A., GONCALVES, A., OXLEY, D., MOHAMMED, H., TURRO, E. & HEMBERGER, M. 2015a. Fgf and Esrrb integrate epigenetic and transcriptional networks that regulate self-renewal of trophoblast stem cells. *Nat Commun,* 6**,** 7776.

LATOS, P. A. & HEMBERGER, M. 2016. From the stem of the placental tree: trophoblast stem cells and their progeny. *Development,* 143**,** 3650-3660.

LATOS, P. A., SIENERTH, A. R., MURRAY, A., SENNER, C. E., MUTO, M., IKAWA, M., OXLEY, D., BURGE, S., COX, B. J. & HEMBERGER, M. 2015b. Elf5-centered transcription factor hub controls trophoblast stem cell self-renewal and differentiation through stoichiometry-sensitive shifts in target gene networks. *Genes Dev,* 29**,** 2435-48.

LEE, B. K., UPRETY, N., JANG, Y. J., TUCKER, S. K., RHEE, C., LEBLANC, L., BECK, S. & KIM, J. 2018. Fosl1 overexpression directly activates trophoblast-specific gene expression programs in embryonic stem cells. *Stem Cell Res,* 26**,** 95-102.

LEE, T. I., JENNER, R. G., BOYER, L. A., GUENTHER, M. G., LEVINE, S. S., KUMAR, R. M., CHEVALIER, B., JOHNSTONE, S. E., COLE, M. F., ISONO, K., KOSEKI, H., FUCHIKAMI, T., ABE, K., MURRAY, H. L., ZUCKER, J. P., YUAN, B., BELL, G. W., HERBOLSHEIMER, E., HANNETT, N. M., SUN, K., ODOM, D. T., OTTE, A. P., VOLKERT, T. L., BARTEL, D. P., MELTON, D. A., GIFFORD, D. K., JAENISCH, R. & YOUNG, R. A. 2006. Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell,* 125**,** 301-13.

LEUNG, C. Y. & ZERNICKA-GOETZ, M. 2013. Angiomotin prevents pluripotent lineage differentiation in mouse embryos via Hippo pathway-dependent and -independent mechanisms. *Nat Commun,* 4**,** 2251.

LI, B. & DEWEY, C. N. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics,* 12**,** 323.

LI, H. & DURBIN, R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics,* 26**,** 589-95.

LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G. & DURBIN, R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics,* 25**,** 2078-9.

LI, Y., MORETTO-ZITA, M., SONCIN, F., WAKELAND, A., WOLFE, L., LEON-GARCIA, S., PANDIAN, R., PIZZO, D., CUI, L., NAZOR, K., LORING, J. F., CRUM, C. P., LAURENT, L. C. & PARAST, M. M. 2013. BMP4-directed trophoblast differentiation of human embryonic stem cells is mediated through a DeltaNp63+ cytotrophoblast stem cell state. *Development,* 140**,** 3965-76.

LINDEBOOM, R. G., SUPEK, F. & LEHNER, B. 2016. The rules and impact of nonsense-mediated mRNA decay in human cancers. *Nat Genet,* 48**,** 1112-8.

LIVAK, K. J. & SCHMITTGEN, T. D. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods,* 25**,** 402-8.

LOVE, M. I., HUBER, W. & ANDERS, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol,* 15**,** 550.

LU, C. W., YABUUCHI, A., CHEN, L., VISWANATHAN, S., KIM, K. & DALEY, G. Q. 2008. Ras-MAPK signaling promotes trophectoderm formation from embryonic stem cells and mouse embryos. *Nat Genet,* 40**,** 921-6.

LUO, J., SLADEK, R., BADER, J. A., MATTHYSSEN, A., ROSSANT, J. & GIGUERE, V. 1997. Placental abnormalities in mouse embryos lacking the orphan nuclear receptor ERR-beta. *Nature,* 388**,** 778-82.

MA, G. T. & LINZER, D. I. 2000. GATA-2 restricts prolactin-like protein A expression to secondary trophoblast giant cells in the mouse. *Biol Reprod,* 63**,** 570-4.

MA, G. T., ROTH, M. E., GROSKOPF, J. C., TSAI, F. Y., ORKIN, S. H., GROSVELD, F., ENGEL, J. D. & LINZER, D. I. 1997. GATA-2 and GATA-3 regulate trophoblast-specific gene expression in vivo. *Development,* 124**,** 907-14.

MAHONY, S., EDWARDS, M. D., MAZZONI, E. O., SHERWOOD, R. I., KAKUMANU, A., MORRISON, C. A., WICHTERLE, H. & GIFFORD, D. K. 2014. An integrated model of multiple-condition ChIP-Seq data reveals predeterminants of Cdx2 binding. *PLoS Comput Biol,* 10**,** e1003501.

MAITRE, J. L., NIWAYAMA, R., TURLIER, H., NEDELEC, F. & HIIRAGI, T. 2015. Pulsatile cell-autonomous contractility drives compaction in the mouse embryo. *Nat Cell Biol,* 17**,** 849-55.

MALASSINE, A., FRENDO, J. L., BLAISE, S., HANDSCHUH, K., GERBAUD, P., TSATSARIS, V., HEIDMANN, T. & EVAIN-BRION, D. 2008. Human endogenous retrovirus-FRD envelope protein (syncytin 2) expression in normal and trisomy 21-affected placenta. *Retrovirology,* 5**,** 6.

MALTEPE, E., KRAMPITZ, G. W., OKAZAKI, K. M., RED-HORSE, K., MAK, W., SIMON, M. C. & FISHER, S. J. 2005. Hypoxia-inducible factor-dependent histone deacetylase activity determines stem cell fate in the placenta. *Development,* 132**,** 3393-403.

MARTELLO, G., SUGIMOTO, T., DIAMANTI, E., JOSHI, A., HANNAH, R., OHTSUKA, S., GOTTGENS, B., NIWA, H. & SMITH, A. 2012. Esrrb is a pivotal target of the Gsk3/Tcf3 axis regulating embryonic stem cell self-renewal. *Cell Stem Cell,* 11**,** 491-504.

MARTIN, G. R. 1981. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci U S A,* 78**,** 7634-8.

MARTIN, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011,* 17**,** 3.

MASUI, S., NAKATAKE, Y., TOYOOKA, Y., SHIMOSATO, D., YAGI, R., TAKAHASHI, K., OKOCHI, H., OKUDA, A., MATOBA, R., SHAROV, A. A., KO, M. S. & NIWA, H. 2007. Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nat Cell Biol,* 9**,** 625-35.

MATHELIER, A., FORNES, O., ARENILLAS, D. J., CHEN, C. Y., DENAY, G., LEE, J., SHI, W., SHYR, C., TAN, G., WORSLEY-HUNT, R., ZHANG, A. W., PARCY, F., LENHARD, B., SANDELIN, A. & WASSERMAN, W. W. 2016. JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res,* 44**,** D110-5.

MCLEAN, C. Y., BRISTOR, D., HILLER, M., CLARKE, S. L., SCHAAR, B. T., LOWE, C. B., WENGER, A. M. & BEJERANO, G. 2010. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol,* 28**,** 495-501.

MESS, A. M. & FERNER, K. J. 2010. Evolution and development of gas exchange structures in Mammalia: the placenta and the lung. *Respir Physiol Neurobiol,* 173 Suppl**,** S74-82.

METZIS, V., STEINHAUSER, S., PAKANAVICIUS, E., GOUTI, M., STAMATAKI, D., IVANOVITCH, K., WATSON, T., RAYON, T., MOUSAVY GHARAVY, S. N., LOVELL-BADGE, R., LUSCOMBE, N. M. & BRISCOE, J. 2018. Nervous System Regionalization Entails Axial Allocation before Neural Differentiation. *Cell,* 175**,** 1105-1118.e17.

MEYER, B. I. & GRUSS, P. 1993. Mouse Cdx-1 expression during gastrulation. *Development,* 117**,** 191-203.

MO, F. E., MUNTEAN, A. G., CHEN, C. C., STOLZ, D. B., WATKINS, S. C. & LAU, L. F. 2002. CYR61 (CCN1) is essential for placental development and vascular integrity. *Mol Cell Biol,* 22**,** 8709-20.

MORAN, D. M., SHEN, H. & MAKI, C. G. 2009. Puromycin-based vectors promote a ROS-dependent recruitment of PML to nuclear inclusions enriched with HSP70 and Proteasomes. *BMC Cell Biol,* 10**,** 32.

MOU, H., SMITH, J. L., PENG, L., YIN, H., MOORE, J., ZHANG, X. O., SONG, C. Q., SHEEL, A., WU, Q., OZATA, D. M., LI, Y., ANDERSON, D. G., EMERSON, C. P., SONTHEIMER, E. J., MOORE, M. J., WENG, Z. & XUE, W. 2017. CRISPR/Cas9-mediated genome editing induces exon skipping by alternative splicing or exon deletion. *Genome Biol,* 18**,** 108.

MOULD, A., MORGAN, M. A., LI, L., BIKOFF, E. K. & ROBERTSON, E. J. 2012. Blimp1/Prdm1 governs terminal differentiation of endovascular trophoblast giant cells and defines multipotent progenitors in the developing placenta. *Genes Dev,* 26**,** 2063-74.

MUROHASHI, M., NAKAMURA, T., TANAKA, S., ICHISE, T., YOSHIDA, N., YAMAMOTO, T., SHIBUYA, M., SCHLESSINGER, J. & GOTOH, N. 2010. An

FGF4-FRS2alpha-Cdx2 axis in trophoblast stem cells induces Bmp4 to regulate proper growth of early mouse embryos. *Stem Cells,* 28**,** 113-21.

MURRAY, A., SIENERTH, A. R. & HEMBERGER, M. 2016. Plet1 is an epigenetically regulated cell surface protein that provides essential cues to direct trophoblast stem cell differentiation. *Sci Rep,* 6**,** 25112.

NADRA, K., ANGHEL, S. I., JOYE, E., TAN, N. S., BASU-MODAK, S., TRONO, D., WAHLI, W. & DESVERGNE, B. 2006. Differentiation of trophoblast giant cells and their metabolic functions are dependent on peroxisome proliferator-activated receptor beta/delta. *Mol Cell Biol,* 26**,** 3266-81.

NAKAYAMA, H., SCOTT, I. C. & CROSS, J. C. 1998. The transition to endoreduplication in trophoblast giant cells is regulated by the mSNA zinc finger transcription factor. *Dev Biol,* 199**,** 150-63.

NAKAYAMA, J., RICE, J. C., STRAHL, B. D., ALLIS, C. D. & GREWAL, S. I. 2001. Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science,* 292**,** 110-3.

NATALE, D. R., HEMBERGER, M., HUGHES, M. & CROSS, J. C. 2009. Activin promotes differentiation of cultured mouse trophoblast stem cells towards a labyrinth cell fate. *Dev Biol,* 335**,** 120-31.

NAU, H. 2001. Teratogenicity of isotretinoin revisited: species variation and the role of all-trans-retinoic acid. *J Am Acad Dermatol,* 45**,** S183-7.

NELSON, A. C., MOULD, A. W., BIKOFF, E. K. & ROBERTSON, E. J. 2017. Mapping the chromatin landscape and Blimp1 transcriptional targets that regulate trophoblast differentiation. *Sci Rep,* 7**,** 6793.

NG, R. K., DEAN, W., DAWSON, C., LUCIFERO, D., MADEJA, Z., REIK, W. & HEMBERGER, M. 2008. Epigenetic restriction of embryonic cell lineage fate by methylation of Elf5. *Nat Cell Biol,* 10**,** 1280-90.

NG, Y. K., GEORGE, K. M., ENGEL, J. D. & LINZER, D. I. 1994. GATA factor activity is required for the trophoblast-specific transcriptional regulation of the mouse placental lactogen I gene. *Development,* 120**,** 3257-66.

NIAKAN, K. K. & EGGAN, K. 2013. Analysis of human embryos from zygote to blastocyst reveals distinct gene expression patterns relative to the mouse. *Dev Biol,* 375**,** 54-64.

NICHOLS, J., ZEVNIK, B., ANASTASSIADIS, K., NIWA, H., KLEWE-NEBENIUS, D., CHAMBERS, I., SCHOLER, H. & SMITH, A. 1998. Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell,* 95**,** 379-91.

NIE, G., LI, Y. & SALAMONSEN, L. A. 2005. Serine protease HtrA1 is developmentally regulated in trophoblast and uterine decidual cells during placental formation in the mouse. *Dev Dyn,* 233**,** 1102-9.

NIKAS, G., AO, A., WINSTON, R. M. & HANDYSIDE, A. H. 1996. Compaction and surface polarity in the human embryo in vitro. *Biol Reprod,* 55**,** 32-7.

NISHIOKA, N., INOUE, K., ADACHI, K., KIYONARI, H., OTA, M., RALSTON, A., YABUTA, N., HIRAHARA, S., STEPHENSON, R. O., OGONUKI, N., MAKITA, R., KURIHARA, H., MORIN-KENSICKI, E. M., NOJIMA, H., ROSSANT, J., NAKAO, K., NIWA, H. & SASAKI, H. 2009. The Hippo signaling pathway components Lats and Yap pattern Tead4 activity to distinguish mouse trophectoderm from inner cell mass. *Dev Cell,* 16**,** 398-410.

NISHIOKA, N., YAMAMOTO, S., KIYONARI, H., SATO, H., SAWADA, A., OTA, M., NAKAO, K. & SASAKI, H. 2008. Tead4 is required for specification of trophectoderm in pre-implantation mouse embryos. *Mech Dev,* 125**,** 270-83.

NISHIYAMA, A., XIN, L., SHAROV, A. A., THOMAS, M., MOWRER, G., MEYERS, E., PIAO, Y., MEHTA, S., YEE, S., NAKATAKE, Y., STAGG, C., SHAROVA, L., CORREA-CERRO, L. S., BASSEY, U., HOANG, H., KIM, E., TAPNIO, R.,

QIAN, Y., DUDEKULA, D., ZALZMAN, M., LI, M., FALCO, G., YANG, H. T., LEE, S. L., MONTI, M., STANGHELLINI, I., ISLAM, M. N., NAGARAJA, R., GOLDBERG, I., WANG, W., LONGO, D. L., SCHLESSINGER, D. & KO, M. S. 2009. Uncovering early response of gene regulatory networks in ESCs by systematic induction of transcription factors. *Cell Stem Cell,* 5**,** 420-33.

NIWA, H., MIYAZAKI, J. & SMITH, A. G. 2000. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat Genet,* 24**,** 372-6.

NIWA, H., TOYOOKA, Y., SHIMOSATO, D., STRUMPF, D., TAKAHASHI, K., YAGI, R. & ROSSANT, J. 2005. Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. *Cell,* 123**,** 917-29.

NORA, E. P., LAJOIE, B. R., SCHULZ, E. G., GIORGETTI, L., OKAMOTO, I., SERVANT, N., PIOLOT, T., VAN BERKUM, N. L., MEISIG, J., SEDAT, J., GRIBNAU, J., BARILLOT, E., BLUTHGEN, N., DEKKER, J. & HEARD, E. 2012. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature,* 485**,** 381-5.

ODIATIS, C. & GEORGIADES, P. 2010. New insights for Ets2 function in trophoblast using lentivirus-mediated gene knockdown in trophoblast stem cells. *Placenta,* 31**,** 630-40.

OHINATA, Y. & TSUKIYAMA, T. 2014. Establishment of trophoblast stem cells under defined culture conditions in mice. *PLoS One,* 9**,** e107308.

OKAE, H., TOH, H., SATO, T., HIURA, H., TAKAHASHI, S., SHIRANE, K., KABAYAMA, Y., SUYAMA, M., SASAKI, H. & ARIMA, T. 2018. Derivation of Human Trophoblast Stem Cells. *Cell Stem Cell,* 22**,** 50-63.e6.

OKITSU, C. Y., HSIEH, J. C. & HSIEH, C. L. 2010. Transcriptional activity affects the H3K4me3 level and distribution in the coding region. *Mol Cell Biol,* 30**,** 2933-46.

PADDISON, P. J., CAUDY, A. A., BERNSTEIN, E., HANNON, G. J. & CONKLIN, D. S. 2002. Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. *Genes Dev,* 16**,** 948-58.

PANDOLFI, P. P., ROTH, M. E., KARIS, A., LEONARD, M. W., DZIERZAK, E., GROSVELD, F. G., ENGEL, J. D. & LINDENBAUM, M. H. 1995. Targeted disruption of the GATA3 gene causes severe abnormalities in the nervous system and in fetal liver haematopoiesis. *Nat Genet,* 11**,** 40-4.

PARAST, M. M., AEDER, S. & SUTHERLAND, A. E. 2001. Trophoblast giant-cell differentiation involves changes in cytoskeleton and cell motility. *Dev Biol,* 230**,** 43-60.

PARISI, T., BECK, A. R., ROUGIER, N., MCNEIL, T., LUCIAN, L., WERB, Z. & AMATI, B. 2003. Cyclins E1 and E2 are required for endoreplication in placental trophoblast giant cells. *Embo j,* 22**,** 4794-803.

PARK, J., LIM, K., KIM, J. S. & BAE, S. 2017. Cas-analyzer: an online tool for assessing genome editing results using NGS data. *Bioinformatics,* 33**,** 286-288.

PEARTON, D. J., SMITH, C. S., REDGATE, E., VAN LEEUWEN, J., DONNISON, M. & PFEFFER, P. L. 2014. Elf5 counteracts precocious trophoblast differentiation by maintaining Sox2 and 3 and inhibiting Hand1 expression. *Dev Biol,* 392**,** 344-57.

PEREZ-GARCIA, V., FINEBERG, E., WILSON, R., MURRAY, A., MAZZEO, C. I., TUDOR, C., SIENERTH, A., WHITE, J. K., TUCK, E., RYDER, E. J., GLEESON, D., SIRAGHER, E., WARDLE-JONES, H., STAUDT, N., WALI, N., COLLINS, J., GEYER, S., BUSCH-NENTWICH, E. M., GALLI, A., SMITH, J. C., ROBERTSON, E., ADAMS, D. J., WENINGER, W. J., MOHUN, T. & HEMBERGER, M. 2018. Placentation defects are highly prevalent in embryonic lethal mouse mutants. *Nature,* 555**,** 463-468.

PETERS, A. H., KUBICEK, S., MECHTLER, K., O'SULLIVAN, R. J., DERIJCK, A. A., PEREZ-BURGOS, L., KOHLMAIER, A., OPRAVIL, S., TACHIBANA, M., SHINKAI, Y., MARTENS, J. H. & JENUWEIN, T. 2003. Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. *Mol Cell,* 12**,** 1577-89.

PETROPOULOS, S., EDSGARD, D., REINIUS, B., DENG, Q., PANULA, S. P., CODELUPPI, S., PLAZA REYES, A., LINNARSSON, S., SANDBERG, R. & LANNER, F. 2016. Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos. *Cell,* 165**,** 1012-26.

PIQUE-REGI, R., DEGNER, J. F., PAI, A. A., GAFFNEY, D. J., GILAD, Y. & PRITCHARD, J. K. 2011. Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. *Genome Res,* 21**,** 447-55.

PLATH, K., FANG, J., MLYNARCZYK-EVANS, S. K., CAO, R., WORRINGER, K. A., WANG, H., DE LA CRUZ, C. C., OTTE, A. P., PANNING, B. & ZHANG, Y. 2003. Role of histone H3 lysine 27 methylation in X inactivation. *Science,* 300**,** 131-5.

PLUSA, B., FRANKENBERG, S., CHALMERS, A., HADJANTONAKIS, A. K., MOORE, C. A., PAPALOPULU, N., PAPAIOANNOU, V. E., GLOVER, D. M. & ZERNICKA-GOETZ, M. 2005. Downregulation of Par3 and aPKC function directs cells towards the ICM in the preimplantation mouse embryo. *J Cell Sci,* 118**,** 505-15.

POLYDOROU, C. & GEORGIADES, P. 2013. Ets2-dependent trophoblast signalling is required for gastrulation progression after primitive streak initiation. *Nat Commun,* 4**,** 1658.

POSFAI, E., PETROPOULOS, S., DE BARROS, F. R., SCHELL, J. P., JURISICA, I., SANDBERG, R., LANNER, F. & ROSSANT, J. 2017. Position- and Hippo signaling-dependent plasticity during lineage segregation in the early mouse embryo. *Elife,* 6.

QUINLAN, A. R. & HALL, I. M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics,* 26**,** 841-2.

RADA-IGLESIAS, A., BAJPAI, R., SWIGUT, T., BRUGMANN, S. A., FLYNN, R. A. & WYSOCKA, J. 2011. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature,* 470**,** 279-83.

RAI, A. & CROSS, J. C. 2014. Development of the hemochorial maternal vascular spaces in the placenta through endothelial and vasculogenic mimicry. *Dev Biol,* 387**,** 131-41.

RAI, A. & CROSS, J. C. 2015. Three-dimensional cultures of trophoblast stem cells autonomously develop vascular-like spaces lined by trophoblast giant cells. *Dev Biol,* 398**,** 110-9.

RAISNER, R., KHARBANDA, S., JIN, L., JENG, E., CHAN, E., MERCHANT, M., HAVERTY, P. M., BAINER, R., CHEUNG, T., ARNOTT, D., FLYNN, E. M., ROMERO, F. A., MAGNUSON, S. & GASCOIGNE, K. E. 2018. Enhancer Activity Requires CBP/P300 Bromodomain-Dependent Histone H3K27 Acetylation. *Cell Rep,* 24**,** 1722-1729.

RALSTON, A., COX, B. J., NISHIOKA, N., SASAKI, H., CHEA, E., RUGG-GUNN, P., GUO, G., ROBSON, P., DRAPER, J. S. & ROSSANT, J. 2010. Gata3 regulates trophoblast development downstream of Tead4 and in parallel to Cdx2. *Development,* 137**,** 395-403.

RALSTON, A. & ROSSANT, J. 2008. Cdx2 acts downstream of cell polarization to cell-autonomously promote trophectoderm fate in the early mouse embryo. *Dev Biol,* 313**,** 614-29.

RAMIREZ, F., RYAN, D. P., GRUNING, B., BHARDWAJ, V., KILPERT, F., RICHTER, A. S., HEYNE, S., DUNDAR, F. & MANKE, T. 2016. deepTools2: a next

generation web server for deep-sequencing data analysis. *Nucleic Acids Res,* 44**,** W160-5.

RAN, F. A., HSU, P. D., WRIGHT, J., AGARWALA, V., SCOTT, D. A. & ZHANG, F. 2013. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc,* 8**,** 2281-2308.

RAWN, S. M. & CROSS, J. C. 2008. The evolution, regulation, and function of placenta-specific genes. *Annu Rev Cell Dev Biol,* 24**,** 159-81.

RAWN, S. M., HUANG, C., HUGHES, M., SHAYKHUTDINOV, R., VOGEL, H. J. & CROSS, J. C. 2015. Pregnancy Hyperglycemia in Prolactin Receptor Mutant, but Not Prolactin Mutant, Mice and Feeding-Responsive Regulation of Placental Lactogen Genes Implies Placental Control of Maternal Glucose Homeostasis. *Biol Reprod,* 93**,** 75.

RAY, S., DUTTA, D., RUMI, M. A., KENT, L. N., SOARES, M. J. & PAUL, S. 2009. Context-dependent function of regulatory elements and a switch in chromatin occupancy between GATA3 and GATA2 regulate Gata2 transcription during trophoblast differentiation. *J Biol Chem,* 284**,** 4978-88.

RAYON, T., MENCHERO, S., NIETO, A., XENOPOULOS, P., CRESPO, M., COCKBURN, K., CANON, S., SASAKI, H., HADJANTONAKIS, A. K., DE LA POMPA, J. L., ROSSANT, J. & MANZANARES, M. 2014. Notch and hippo converge on Cdx2 to specify the trophectoderm lineage in the mouse blastocyst. *Dev Cell,* 30**,** 410-22.

RAYON, T., MENCHERO, S., ROLLAN, I., ORS, I., HELNESS, A., CRESPO, M., NIETO, A., AZUARA, V., ROSSANT, J. & MANZANARES, M. 2016. Distinct mechanisms regulate Cdx2 expression in the blastocyst and in trophoblast stem cells. *Sci Rep,* 6**,** 27139.

REA, S., EISENHABER, F., O'CARROLL, D., STRAHL, B. D., SUN, Z. W., SCHMID, M., OPRAVIL, S., MECHTLER, K., PONTING, C. P., ALLIS, C. D. & JENUWEIN, T. 2000. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature,* 406**,** 593-9.

RENAUD, S. J., KUBOTA, K., RUMI, M. A. & SOARES, M. J. 2014. The FOS transcription factor family differentially controls trophoblast migration and invasion. *J Biol Chem,* 289**,** 5025-39.

RHEE, C., LEE, B. K., BECK, S., ANJUM, A., COOK, K. R., POPOWSKI, M., TUCKER, H. O. & KIM, J. 2014. Arid3a is essential to execution of the first cell fate decision via direct embryonic and extraembryonic transcriptional regulation. *Genes Dev,* 28**,** 2219-32.

RHEE, C., LEE, B. K., BECK, S., LEBLANC, L., TUCKER, H. O. & KIM, J. 2017. Mechanisms of transcription factor-mediated direct reprogramming of mouse embryonic stem cells to trophoblast stem-like cells. *Nucleic Acids Res,* 45**,** 10103-10114.

RILEY, P., ANSON-CARTWRIGHT, L. & CROSS, J. C. 1998. The Hand1 bHLH transcription factor is essential for placentation and cardiac morphogenesis. *Nat Genet,* 18**,** 271-5.

ROBERTS, R. M., LOH, K. M., AMITA, M., BERNARDO, A. S., ADACHI, K., ALEXENKO, A. P., SCHUST, D. J., SCHULZ, L. C., TELUGU, B. P., EZASHI, T. & PEDERSEN, R. A. 2014. Differentiation of trophoblast cells from human embryonic stem cells: to be or not to be? *Reproduction,* 147**,** D1-12.

ROBINSON, J. T., THORVALDSDOTTIR, H., WINCKLER, W., GUTTMAN, M., LANDER, E. S., GETZ, G. & MESIROV, J. P. 2011. Integrative genomics viewer. *Nat Biotechnol,* 29**,** 24-6.

ROSSANT, J. 2015. Mouse and human blastocyst-derived stem cells: vive les differences. *Development,* 142**,** 9-12.

ROSSANT, J. & CROSS, J. C. 2001. Placental development: lessons from mouse mutants. *Nat Rev Genet,* 2**,** 538-48.

ROSSANT, J., GUILLEMOT, F., TANAKA, M., LATHAM, K., GERTENSTEIN, M. & NAGY, A. 1998. Mash2 is expressed in oogenesis and preimplantation development but is not required for blastocyst formation. *Mech Dev,* 73**,** 183-91.

ROSSI, A., KONTARAKIS, Z., GERRI, C., NOLTE, H., HOLPER, S., KRUGER, M. & STAINIER, D. Y. 2015. Genetic compensation induced by deleterious mutations but not gene knockdowns. *Nature,* 524**,** 230-3.

ROUSSEEUW, P. J. & RUTS, I. 1997. *"The Bagplot: A Bivariate Box-and-Whiskers Plot,"*.

RUGG-GUNN, P. J., COX, B. J., RALSTON, A. & ROSSANT, J. 2010. Distinct histone modifications in stem cell lines and tissue lineages from the early mouse embryo. *Proc Natl Acad Sci U S A,* 107**,** 10783-90.

RUSS, A. P., WATTLER, S., COLLEDGE, W. H., APARICIO, S. A., CARLTON, M. B., PEARCE, J. J., BARTON, S. C., SURANI, M. A., RYAN, K., NEHLS, M. C., WILSON, V. & EVANS, M. J. 2000. Eomesodermin is required for mouse trophoblast development and mesoderm formation. *Nature,* 404**,** 95-9.

SABA-EL-LEIL, M. K., VELLA, F. D., VERNAY, B., VOISIN, L., CHEN, L., LABRECQUE, N., ANG, S. L. & MELOCHE, S. 2003. An essential function of the mitogen-activated protein kinase Erk2 in mouse trophoblast development. *EMBO Rep,* 4**,** 964-8.

SAN ROMAN, A. K., TOVAGLIERI, A., BREAULT, D. T. & SHIVDASANI, R. A. 2015. Distinct Processes and Transcriptional Targets Underlie CDX2 Requirements in Intestinal Stem Cells and Differentiated Villus Cells. *Stem Cell Reports,* 5**,** 673-81.

SANTOS-ROSA, H., SCHNEIDER, R., BANNISTER, A. J., SHERRIFF, J., BERNSTEIN, B. E., EMRE, N. C., SCHREIBER, S. L., MELLOR, J. & KOUZARIDES, T. 2002. Active genes are tri-methylated at K4 of histone H3. *Nature,* 419**,** 407-11.

SATI, S. & CAVALLI, G. 2017. Chromosome conformation capture technologies and their impact in understanding genome function. *Chromosoma,* 126**,** 33-44.

SCHENKE-LAYLAND, K., ANGELIS, E., RHODES, K. E., HEYDARKHAN-HAGVALL, S., MIKKOLA, H. K. & MACLELLAN, W. R. 2007. Collagen IV induces trophoectoderm differentiation of mouse embryonic stem cells. *Stem Cells,* 25**,** 1529-38.

SCHMIDT, A., MORALES-PRIETO, D. M., PASTUSCHEK, J., FROHLICH, K. & MARKERT, U. R. 2015. Only humans have human placentas: molecular differences between mice and humans. *J Reprod Immunol,* 108**,** 65-71.

SCOTT, I. C., ANSON-CARTWRIGHT, L., RILEY, P., REDA, D. & CROSS, J. C. 2000. The HAND1 basic helix-loop-helix transcription factor regulates trophoblast differentiation via multiple mechanisms. *Mol Cell Biol,* 20**,** 530-41.

SENNER, C. E., KRUEGER, F., OXLEY, D., ANDREWS, S. & HEMBERGER, M. 2012. DNA methylation profiles define stem cell identity and reveal a tight embryonic-extraembryonic lineage boundary. *Stem Cells,* 30**,** 2732-45.

SHAULIAN, E. & KARIN, M. 2002. AP-1 as a regulator of cell life and death. *Nat Cell Biol,* 4**,** E131-6.

SHEN, L., SHAO, N. Y., LIU, X., MAZE, I., FENG, J. & NESTLER, E. J. 2013. diffReps: detecting differential chromatin modification sites from ChIP-seq data with biological replicates. *PLoS One,* 8**,** e65598.

SILVA, J., MAK, W., ZVETKOVA, I., APPANAH, R., NESTEROVA, T. B., WEBSTER, Z., PETERS, A. H., JENUWEIN, T., OTTE, A. P. & BROCKDORFF, N. 2003. Establishment of histone h3 methylation on the inactive X chromosome requires

transient recruitment of Eed-Enx1 polycomb group complexes. *Dev Cell,* 4**,** 481-95.

SILVA, J. F. & SERAKIDES, R. 2016. Intrauterine trophoblast migration: A comparative view of humans and rodents. *Cell Adh Migr,* 10**,** 88-110.

SIMMONS, D. G. & CROSS, J. C. 2005. Determinants of trophoblast lineage and cell subtype specification in the mouse placenta. *Dev Biol,* 284**,** 12-24.

SIMMONS, D. G., FORTIER, A. L. & CROSS, J. C. 2007. Diverse subtypes and developmental origins of trophoblast giant cells in the mouse placenta. *Dev Biol,* 304**,** 567-78.

SIMMONS, D. G., NATALE, D. R., BEGAY, V., HUGHES, M., LEUTZ, A. & CROSS, J. C. 2008. Early patterning of the chorion leads to the trilaminar trophoblast cell structure in the placental labyrinth. *Development,* 135**,** 2083-91.

SMITH, P. J., ZHANG, C., WANG, J., CHEW, S. L., ZHANG, M. Q. & KRAINER, A. R. 2006. An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers. *Hum Mol Genet,* 15**,** 2490-508.

SOLOMON, M. J., LARSEN, P. L. & VARSHAVSKY, A. 1988. Mapping protein-DNA interactions in vivo with formaldehyde: evidence that histone H4 is retained on a highly transcribed gene. *Cell,* 53**,** 937-47.

SONCIN, F., KHATER, M., TO, C., PIZZO, D., FARAH, O., WAKELAND, A., ARUL NAMBI RAJAN, K., NELSON, K. K., CHANG, C. W., MORETTO-ZITA, M., NATALE, D. R., LAURENT, L. C. & PARAST, M. M. 2018. Comparative analysis of mouse and human placentae across gestation reveals species-specific regulators of placental development. *Development,* 145.

SONCIN, F., NATALE, D. & PARAST, M. M. 2015. Signaling pathways in mouse and human trophoblast differentiation: a comparative review. *Cell Mol Life Sci,* 72**,** 1291-302.

STEPTOE, P. C., EDWARDS, R. G. & PURDY, J. M. 1971. Human blastocysts grown in culture. *Nature,* 229**,** 132-3.

STRUMPF, D., MAO, C. A., YAMANAKA, Y., RALSTON, A., CHAWENGSAKSOPHAK, K., BECK, F. & ROSSANT, J. 2005. Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. *Development,* 132**,** 2093-102.

SUBRAMANIAN, V., MEYER, B. I. & GRUSS, P. 1995. Disruption of the murine homeobox gene Cdx1 affects axial skeletal identities by altering the mesodermal expression domains of Hox genes. *Cell,* 83**,** 641-53.

SUNG, M. H., GUERTIN, M. J., BAEK, S. & HAGER, G. L. 2014. DNase footprint signatures are dictated by factor dynamics and DNA sequence. *Mol Cell,* 56**,** 275-285.

SWINSTEAD, E. E., MIRANDA, T. B., PAAKINAHO, V., BAEK, S., GOLDSTEIN, I., HAWKINS, M., KARPOVA, T. S., BALL, D., MAZZA, D., LAVIS, L. D., GRIMM, J. B., MORISAKI, T., GRONTVED, L., PRESMAN, D. M. & HAGER, G. L. 2016. Steroid Receptors Reprogram FoxA1 Occupancy through Dynamic Chromatin Transitions. *Cell,* 165**,** 593-605.

SZABO, Q., BANTIGNIES, F. & CAVALLI, G. 2019. Principles of genome folding into topologically associating domains. *Sci Adv,* 5**,** eaaw1668.

TABANSKY, I., LENARCIC, A., DRAFT, R. W., LOULIER, K., KESKIN, D. B., ROSAINS, J., RIVERA-FELICIANO, J., LICHTMAN, J. W., LIVET, J., STERN, J. N., SANES, J. R. & EGGAN, K. 2013. Developmental bias in cleavage-stage mouse blastomeres. *Curr Biol,* 23**,** 21-31.

TANAKA, S., KUNATH, T., HADJANTONAKIS, A. K., NAGY, A. & ROSSANT, J. 1998. Promotion of trophoblast stem cell proliferation by FGF4. *Science,* 282**,** 2072-5.

TESSER, R. B., SCHERHOLZ, P. L., DO NASCIMENTO, L. & KATZ, S. G. 2010. Trophoblast glycogen cells differentiate early in the mouse ectoplacental cone: putative role during placentation. *Histochem Cell Biol,* 134**,** 83-92.

THORVALDSDOTTIR, H., ROBINSON, J. T. & MESIROV, J. P. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform,* 14**,** 178-92.

TREMBLAY, G. B., KUNATH, T., BERGERON, D., LAPOINTE, L., CHAMPIGNY, C., BADER, J. A., ROSSANT, J. & GIGUERE, V. 2001. Diethylstilbestrol regulates trophoblast stem cell differentiation as a ligand of orphan nuclear receptor ERR beta. *Genes Dev,* 15**,** 833-8.

TSAKIRIDIS, A., HUANG, Y., BLIN, G., SKYLAKI, S., WYMEERSCH, F., OSORNO, R., ECONOMOU, C., KARAGIANNI, E., ZHAO, S., LOWELL, S. & WILSON, V. 2014. Distinct Wnt-driven primitive streak-like populations reflect in vivo lineage precursors. *Development,* 141**,** 1209-21.

UNTERGASSER, A., NIJVEEN, H., RAO, X., BISSELING, T., GEURTS, R. & LEUNISSEN, J. A. 2007. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res,* 35**,** W71-4.

UY, G. D., DOWNS, K. M. & GARDNER, R. L. 2002. Inhibition of trophoblast stem cell potential in chorionic ectoderm coincides with occlusion of the ectoplacental cavity in the mouse. *Development,* 129**,** 3913-24.

VAN DE VEN, C., BIALECKA, M., NEIJTS, R., YOUNG, T., ROWLAND, J. E., STRINGER, E. J., VAN ROOIJEN, C., MEIJLINK, F., NOVOA, A., FREUND, J. N., MALLO, M., BECK, F. & DESCHAMPS, J. 2011. Concerted involvement of Cdx/Hox genes and Wnt signaling in morphogenesis of the caudal neural tube and cloacal derivatives from the posterior growth zone. *Development,* 138**,** 3451-62.

VAN DEN BERG, D. L., ZHANG, W., YATES, A., ENGELEN, E., TAKACS, K., BEZSTAROSTI, K., DEMMERS, J., CHAMBERS, I. & POOT, R. A. 2008. Estrogen-related receptor beta interacts with Oct4 to positively regulate Nanog gene expression. *Mol Cell Biol,* 28**,** 5986-95.

VAN NES, J., DE GRAAFF, W., LEBRIN, F., GERHARD, M., BECK, F. & DESCHAMPS, J. 2006. The Cdx4 mutation affects axial development and reveals an essential role of Cdx genes in the ontogenesis of the placental labyrinth in mice. *Development,* 133**,** 419-28.

VAN ROOIJEN, C., SIMMINI, S., BIALECKA, M., NEIJTS, R., VAN DE VEN, C., BECK, F. & DESCHAMPS, J. 2012. Evolutionarily conserved requirement of Cdx for post-occipital tissue emergence. *Development,* 139**,** 2576-83.

VERZI, M. P., SHIN, H., HE, H. H., SULAHIAN, R., MEYER, C. A., MONTGOMERY, R. K., FLEET, J. C., BROWN, M., LIU, X. S. & SHIVDASANI, R. A. 2010. Differentiation-specific histone modifications reveal dynamic chromatin interactions and partners for the intestinal transcription factor CDX2. *Dev Cell,* 19**,** 713-26.

VERZI, M. P., SHIN, H., SAN ROMAN, A. K., LIU, X. S. & SHIVDASANI, R. A. 2013. Intestinal master transcription factor CDX2 controls chromatin access for partner transcription factor binding. *Mol Cell Biol,* 33**,** 281-92.

VINOT, S., LE, T., OHNO, S., PAWSON, T., MARO, B. & LOUVET-VALLEE, S. 2005. Asymmetric distribution of PAR proteins in the mouse embryo begins at the 8-cell stage during compaction. *Dev Biol,* 282**,** 307-19.

WAGIH, O. 2017. ggseqlogo: a versatile R package for drawing sequence logos. *Bioinformatics,* 33**,** 3645-3647.

WAMAITHA, S. E. & NIAKAN, K. K. 2018. Human Pre-gastrulation Development. *Curr Top Dev Biol,* 128**,** 295-338.

WANG, H., XIE, H., SUN, X., TRANGUCH, S., ZHANG, H., JIA, X., WANG, D., DAS, S. K., DESVERGNE, B., WAHLI, W., DUBOIS, R. N. & DEY, S. K. 2007. Stage-specific integration of maternal and embryonic peroxisome proliferator-activated receptor delta signaling is critical to pregnancy success. *J Biol Chem,* 282**,** 37770-82.

WANG, Y., SONG, F., ZHANG, B., ZHANG, L., XU, J., KUANG, D., LI, D., CHOUDHARY, M. N. K., LI, Y., HU, M., HARDISON, R., WANG, T. & YUE, F. 2018. The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. *Genome Biol,* 19**,** 151.

WATANABE, Y., MIYASAKA, K. Y., KUBO, A., KIDA, Y. S., NAKAGAWA, O., HIRATE, Y., SASAKI, H. & OGURA, T. 2017. Notch and Hippo signaling converge on Strawberry Notch 1 (Sbno1) to synergistically activate Cdx2 during specification of the trophectoderm. *Sci Rep,* 7**,** 46135.

WATSON, E. D. & CROSS, J. C. 2005. Development of structures and transport functions in the mouse placenta. *Physiology (Bethesda),* 20**,** 180-93.

WATSON, E. D., HUGHES, M., SIMMONS, D. G., NATALE, D. R., SUTHERLAND, A. E. & CROSS, J. C. 2011. Cell-cell adhesion defects in Mrj mutant trophoblast cells are associated with failure to pattern the chorion during early placental development. *Dev Dyn,* 240**,** 2505-19.

WEI, S., ZOU, Q., LAI, S., ZHANG, Q., LI, L., YAN, Q., ZHOU, X., ZHONG, H. & LAI, L. 2016. Conversion of embryonic stem cells into extraembryonic lineages by CRISPR-mediated activators. *Sci Rep,* 6**,** 19648.

WELSH, A. O. & ENDERS, A. C. 1987. Trophoblast-decidual cell interactions and establishment of maternal blood circulation in the parietal yolk sac placenta of the rat. *Anat Rec,* 217**,** 203-19.

WEN, F., TYNAN, J. A., CECENA, G., WILLIAMS, R., MUNERA, J., MAVROTHALASSITIS, G. & OSHIMA, R. G. 2007. Ets2 is required for trophoblast stem cell self-renewal. *Dev Biol,* 312**,** 284-99.

WERLING, U. & SCHORLE, H. 2002. Transcription factor gene AP-2 gamma essential for early murine development. *Mol Cell Biol,* 22**,** 3149-56.

WICKLOW, E., BLIJ, S., FRUM, T., HIRATE, Y., LANG, R. A., SASAKI, H. & RALSTON, A. 2014. HIPPO pathway members restrict SOX2 to the inner cell mass where it promotes ICM fates in the mouse blastocyst. *PLoS Genet,* 10**,** e1004618.

WINGER, Q., HUANG, J., AUMAN, H. J., LEWANDOSKI, M. & WILLIAMS, T. 2006. Analysis of transcription factor AP-2 expression and function during mouse preimplantation development. *Biol Reprod,* 75**,** 324-33.

WOODS, L., PEREZ-GARCIA, V. & HEMBERGER, M. 2018. Regulation of Placental Development and Its Impact on Fetal Growth-New Insights From Mouse Models. *Front Endocrinol (Lausanne),* 9**,** 570.

WU, G., GENTILE, L., FUCHIKAMI, T., SUTTER, J., PSATHAKI, K., ESTEVES, T. C., ARAUZO-BRAVO, M. J., ORTMEIER, C., VERBERK, G., ABE, K. & SCHOLER, H. R. 2010. Initiation of trophectoderm lineage specification in mouse embryos is independent of Cdx2. *Development,* 137**,** 4159-69.

WU, T., WANG, H., HE, J., KANG, L., JIANG, Y., LIU, J., ZHANG, Y., KOU, Z., LIU, L., ZHANG, X. & GAO, S. 2011. Reprogramming of trophoblast stem cells into pluripotent stem cells by Oct4. *Stem Cells,* 29**,** 755-63.

WU, Y., LIANG, D., WANG, Y., BAI, M., TANG, W., BAO, S., YAN, Z., LI, D. & LI, J. 2013. Correction of a genetic disease in mouse via use of CRISPR-Cas9. *Cell Stem Cell,* 13**,** 659-62.

XU, R. H., CHEN, X., LI, D. S., LI, R., ADDICKS, G. C., GLENNON, C., ZWAKA, T. P. & THOMSON, J. A. 2002. BMP4 initiates human embryonic stem cell differentiation to trophoblast. *Nat Biotechnol,* 20**,** 1261-4.

YAGI, R., KOHN, M. J., KARAVANOVA, I., KANEKO, K. J., VULLHORST, D., DEPAMPHILIS, M. L. & BUONANNO, A. 2007. Transcription factor TEAD4 specifies the trophectoderm lineage at the beginning of mammalian development. *Development,* 134**,** 3827-36.

YAMADA, K., KANDA, H., TANAKA, S., TAKAMATSU, N., SHIBA, T. & ITO, M. 2006. Sox15 enhances trophoblast giant cell differentiation induced by Hand1 in mouse placenta. *Differentiation,* 74**,** 212-21.

YAMAMOTO, H., FLANNERY, M. L., KUPRIYANOV, S., PEARCE, J., MCKERCHER, S. R., HENKEL, G. W., MAKI, R. A., WERB, Z. & OSHIMA, R. G. 1998. Defective trophoblast function in mice with a targeted mutation of Ets2. *Genes Dev,* 12**,** 1315-26.

YAMANAKA, Y., RALSTON, A., STEPHENSON, R. O. & ROSSANT, J. 2006. Cell and molecular regulation of the mouse blastocyst. *Dev Dyn,* 235**,** 2301-14.

YAN, J., TANAKA, S., ODA, M., MAKINO, T., OHGANE, J. & SHIOTA, K. 2001. Retinoic acid promotes differentiation of trophoblast stem cells to a giant cell fate. *Dev Biol,* 235**,** 422-32.

YANG, H., WANG, H., SHIVALILA, C. S., CHENG, A. W., SHI, L. & JAENISCH, R. 2013. One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. *Cell,* 154**,** 1370-9.

YE, J., COULOURIS, G., ZARETSKAYA, I., CUTCUTACHE, I., ROZEN, S. & MADDEN, T. L. 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics,* 13**,** 134.

YIN, Y., MORGUNOVA, E., JOLMA, A., KAASINEN, E., SAHU, B., KHUND-SAYEED, S., DAS, P. K., KIVIOJA, T., DAVE, K., ZHONG, F., NITTA, K. R., TAIPALE, M., POPOV, A., GINNO, P. A., DOMCKE, S., YAN, J., SCHUBELER, D., VINSON, C. & TAIPALE, J. 2017. Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science,* 356.

YOUNG, T., ROWLAND, J. E., VAN DE VEN, C., BIALECKA, M., NOVOA, A., CARAPUCO, M., VAN NES, J., DE GRAAFF, W., DULUC, I., FREUND, J. N., BECK, F., MALLO, M. & DESCHAMPS, J. 2009. Cdx and Hox genes differentially regulate posterior axial growth in mammalian embryos. *Dev Cell,* 17**,** 516-26.

ZARET, K. S. & CARROLL, J. S. 2011. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev,* 25**,** 2227-41.

ZHANG, Y., LIU, T., MEYER, C. A., EECKHOUTE, J., JOHNSON, D. S., BERNSTEIN, B. E., NUSBAUM, C., MYERS, R. M., BROWN, M., LI, W. & LIU, X. S. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol,* 9**,** R137.

ZHAO, Z., TAVOOSIDANA, G., SJOLINDER, M., GONDOR, A., MARIANO, P., WANG, S., KANDURI, C., LEZCANO, M., SANDHU, K. S., SINGH, U., PANT, V., TIWARI, V., KURUKUTI, S. & OHLSSON, R. 2006. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat Genet,* 38**,** 1341-7.

ZHU, D., HOLZ, S., METZGER, E., PAVLOVIC, M., JANDAUSCH, A., JILG, C., GALGOCZY, P., HERZ, C., MOSER, M., METZGER, D., GUNTHER, T., ARNOLD, S. J. & SCHULE, R. 2014. Lysine-specific demethylase 1 regulates differentiation onset and migration of trophoblast stem cells. *Nat Commun,* 5**,** 3174.